DOCUMENT RESUME

ED 399 304 TM 025 972

AUTHOR Ingels, Steven J.; Quinn, Peggy
TITLE Sample Exclusion in NELS:88. Characteristics of Base Year Ineligible Students: Changes in Eligibility Status after Four Years. Technical Report.
INSTITUTION National Opinion Research Center, Chicago, Ill.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO NCES-96-723
PUB DATE May 96
NOTE 123p.
AVAILABLE FROM National Data Resource Center, 703-845-3151; fax: 703-820-7465 (free).
PUB TYPE Statistical Data (110) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Data Analysis; *Educational Experience; Elementary Secondary Education; *Eligibility; Grade 8; Longitudinal Studies; National Surveys; Probability; *Research Methodology; Sample Size; *Sampling; Tables (Data)
IDENTIFIERS *National Education Longitudinal Study 1988

ABSTRACT

The National Education Longitudinal Studies Program (NELS) is a long-term effort that encompasses the educational experience of youth from three decades. This report is based on the National Education Longitudinal Study of 1988 (NELS:88), a component of the NELS series that began as a national probability sample of eighth grade schools and students in 1988. These students have been followed over time at 2-year intervals, with the most recent followup taking place in 1994. The report is one of a series of methodological monographs that deal with technical issues concerning the analysis of NELS:88 data and the sampling and survey methodology of the study. This methodology monograph addresses issues of student sample exclusion in NELS:88, particularly the problem posed by eighth graders who were declared ineligible for the study owing to factors--physical or mental disabilities, or lack of proficiency in the English language--deemed by their schools to pose a barrier to participation. Chapter 1 explains the exclusion of some students from the survey and discusses the reliability and validity issues exclusion poses. Chapter 2 reviews characteristics of base-year ineligibles, and chapter 3 discusses the eligibility status of the excluded 188 eighth graders in 1992. Chapter 4 considers other potential sources of sample undercoverage, and chapter 5 discusses the need for greater inclusiveness and recommendations to minimize sources of undercoverage. Chapter 6 is a 105-item bibliography. Five appendixes contain three supplementary tables, screening for eligibility reassessment, study eligibility criteria, technical notes, and a glossary. (Contains eight tables, four appendix tables, five figures, and one appendix figure.) (SLD)

7M

ED 399 304

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report **May 1996**

Sample Exclusion in NELS:88

Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

TM025972

U.S. Department of Education
Office of Educational Research and Improvement **NCES 96-723**

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report May 1996

Sample Exclusion in NELS:88

Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years

Steven J. Ingels
National Opinion Research Center (NORC)

Peggy Quinn, Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement NCES 96-723

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Jeanne E. Griffith
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208–5574

May 1996

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status after Four Years* [National Education Longitudinal Study of 1988 Second Follow-Up], NCES 96-723, by Steven J. Ingels. Peggy Quinn, project officer. Washington, D.C.: 1996.

Contact:
Peggy Quinn
(202) 219–1743

For free single copies, call the National Data Resource Center at (703) 845–3151
or send a FAX request to (703) 820–7465.

Foreword

The NCES National Education Longitudinal Studies (NELS) program is a long-term effort that encompasses the educational experience of youth from three decades -- the 1970s, 1980s, and 1990s. The general aim of the NELS program is to study the educational, vocational, and personal development of students at various grade levels, and the personal, familial, social, institutional, and cultural factors that may affect that development. The NELS program currently consists of three major studies: the National Longitudinal Study of the High School Class of 1972 (NLS-72); High School and Beyond (HS&B of 1980); and the National Education Longitudinal Study of 1988 (NELS:88). A fourth study that will begin with a kindergarten cohort -- the Early Childhood Longitudinal Study (ECLS) -- is presently in its design phase. A major purpose of the NELS series is to provide comparative data at different points in time that are germane to educational policy and that permit examination of trends relevant to educational and career development and societal roles.

This report is based on NELS:88, which began as a national probability sample of eighth grade schools and students in the spring of 1988. The NELS:88 eighth graders have been followed over time at two year intervals, with the most recent follow-up taking place in 1994, when most of the cohort had been out of school for two years. A fourth follow-up is scheduled for 1998.

This report is one of a series of methodological monographs that deal with technical issues surrounding the analysis of NELS:88 data and the sampling and survey methodology of the study. While the NELS:88 data set contains a wealth of information that can be used to answer research and policy questions, much too can be learned toward the improvement of future surveys by an examination of the methods used in such a study and their success and shortcomings.

Issues of inclusion and exclusion pose many difficulties that are important to address. Traditionally, a certain proportion of students have been systematically excluded both from cross-sectional assessments such as the National Assessment of Educational Progress (NAEP), and from longitudinal studies such as High School and Beyond (HS&B) and NELS:88. Categories of students have been systematically excluded for one of two reasons: either because of physical or mental disabilities, or because of limited English language proficiency.

There are two principal rationales for such exclusion. The first consideration is that for some students, completing a cognitive test or related survey forms may be an onerous or even an emotionally distressing task. In consequence, some Individualized Education Plans (IEPs) for special education students specify that they should not be tested. A second consideration is that current assessment instruments may not accurately measure the achievement of students with limited English proficiency or severe disabilities, hence might produce results of questionable validity.

On the other hand, systematic exclusion of any group or category of students risks undercoverage bias in the estimates derived from the study. The potential lack of representativeness of a sample that does not include all students both distorts overall results and leaves gaps in our national data for two particular groups of central policy concern -- students with limited English proficiency and students with disabilities. Hence Section 421 (c)(3) of the Perkins Act enjoins the Secretary of Education to ensure that appropriate methodologies are used in assessments of students with limited English proficiency and students with handicaps to ensure valid and reliable comparisons with the general student population and across program areas.

This monograph is intended to contribute to the discussion of issues of surveying and assessing students with disabilities and limited English proficiency by reporting on the experience of NELS:88 with these populations and by making recommendations, drawn in part from the NELS:88 experience, for improved future practice.

Paul Planchon
Associate Commissioner
Elementary and Secondary Education
Statistics Division

Jeffrey A. Owings, Chief
Longitudinal Studies Branch
Elementary and Secondary Education
Statistics Division

Table of Contents

Page

Foreword iii

Acknowledgments viii

Introduction ix

Summary xiii

Chapter 1: Excluded Students In the NELS:88 Base Year 1

1.1 Background 1
1.1.1. Exclusion of Students 1
1.2 Exclusion as an Undercoverage Problem 3
1.3 Reliability and Validity Issues: Were the Base Year Exclusion Criteria Applied Consistently?; Are Excluded Students Unable to Participate? 6
1.4 Studies of the Base Year Ineligibles in NELS:88 First and Second Follow-Ups: Objectives 9
1.5 The Base Year Ineligibles Studies: Implementation 10
1.5.1. Sampling 10
1.5.2. Instrumentation 10
1.5.3. Eligibility Screening 10
1.5.4. Modifications or Accommodations to Include IEP and LEP Students 11

Chapter 2: Characteristics of Base Year Ineligibles 15

2.1 Reason for Exclusion 15
2.2 Sex 19
2.3 Race/ethnicity 21
2.4 1992 Enrollment Status: Students and Dropouts; Grade Progression 23

Chapter 3: Eligibility Status of Excluded 1988 Eighth Graders in 1992 27

Chapter 4: Other Potential Sources of Sample Undercoverage 33

Chapter 5: Conclusions and Recommendations 37

5.1 The Need for Greater Inclusiveness 37
5.2 Recommendations Concerning Greater Inclusiveness 39
5.3 Recommendations Concerning Minimizing Other Sources of Undercoverage 52
5.4 Recommendations with Respect to Sampling Lists 52

Chapter 6: Bibliography 55

List of Appendices

Page

Appendix A: Supplemental Tables A-1

Appendix B: Screener for Eligibility Reassessment B-1

Appendix C: HS&B and NAEP Eligibility Criteria C-1

Appendix D: Technical Notes D-1

Appendix E: Glossary E-1

List of Tables

Table A: Bias Estimates for Eighth Grade Cohort Dropout Rate, 1988-1990. (Percentage of Spring Term 1988 Eighth Graders Not in School Spring Term 1990) 5

Table 1: Characteristics of the Base Year Expanded and Ineligible Samples: as Percentage of Expanded Sample and as Percentage of BYIs 15

Table 2: Public School Children 0 to 21 Years Old Served in Federally Supported Special Education Programs in 1987-88 by Type of Disability 18

Table 3: 1992 School Enrollment Status of the Base Year (1988) Ineligible Sample: Summary 24

Table 4: Expanded Sample 1992 Eligibility Status Summary 28

Table 5: Summary of Final 1992 Statuses for 1988 Excluded Students ("Followed-back" Sample) in Unweighted Percents 29

Table 6: Expanded Sample 1992 Eligibility Status by Type of Exclusion 30

Table 7: Public Elementary School Students in Special Education, 1991-92 48

Table A-1: Characteristics of 1988 Excluded Students by 1992 Eligibility Status A-2

Table A-2: 1988 Exclusion Reason by Race, Sex, Race and Sex, and 1992 Enrollment A-4

Table A-3: 1992 Sample Ns and Status of 1988 Excluded Students by Selected Characteristics A-9

Table C-1: NAEP 1992 Mathematics Assessment IEP/LEP Exclusion Statistics C-6

List of Figures

Page

Figure 1: Base Year Ineligibles by Reason of Exclusion in Weighted Percents 17

Figure 2: Exclusion Status by Gender Breakdown in Weighted Percents 20

Figure 3: Racial/Ethnic Breakdown for Eligibles and BYIs with a Mental Disability in Weighted Percents 22

Figure 4: Base Year (1988) Ineligibles in 1992: Grade Progression by Reason for Exclusion 25

Figure 5: 1992 Eligibility Status of 1988 Excluded Eighth Graders, by Reason for Exclusion 31

Figure D-1: NELS:88 Expanded Cohort, Sample Sizes by Year and Enrollment Status D-4

Acknowledgments

This monograph was prepared by Steven Ingels of the National Opinion Research Center (NORC) at The University of Chicago. The author benefitted from discussions of sample inclusion and related issues with a number of individuals, particularly Don Rock of Educational Testing Service, Norman Bradburn of NORC, and Kevin McGrew, Martha Thurlow, and James Ysseldyke of the National Center on Educational Outcomes. The author is particularly grateful to Jeff Owings of the National Center for Education Statistics for the strong support of efforts to identify and address NELS:88 undercoverage problems that he and his NCES colleagues have offered from the base year of NELS:88 to the present. The author would also like to thank Jules Goodison of Educational Testing Service who, in response to requests for information to assist the planning of the NELS:88 study of base year ineligible students, provided unpublished data on students excluded by NAEP.

Technical review of this document was provided by Thomas B. Hoffer at NORC, and by the following NCES staff members: Peggy Quinn, Ralph Lee, Jeffrey Owings, James Houser, Arnold Goldstein, and Robert Burton. External reviewers were Donald McLaughlin of the American Institutes for Research, and Milagros Lanauze of the Office of Bilingual Education and Minority Languages Affairs. The author wishes to thank each of these individuals for their careful reading of this document and for their valuable suggestions.

Data analysis support was provided by Cecilia Lambert of NORC. Sue Sehra of NORC prepared the expanded sample file, and Jeffrey C. Cothran of NORC produced graphics and assisted in the final processing of this manuscript. Bernard Dugoni helped address adjudication comments in the final draft.

Steven Ingels has served as Project Director for NELS:88 from the Base Year through Second Follow-Up. Other NORC staff who have played a key role in the study include Kathryn L. Dowd, Lisa Thalji, Leslie A. Scott, John R. Taylor, and Martin R. Frankel. Educational Testing Service (ETS) served as the test development subcontractor to NORC. Donald A. Rock directs the ETS component. Peggy Quinn is the NCES NELS:88 Second Follow-Up Project Officer.

Introduction

Overview of NELS:88. The information reported here is taken from the National Education Longitudinal Study of 1988 (NELS:88). The *base year* of NELS:88 was the first stage of a major longitudinal effort designed to provide trend data about critical transitions experienced by students as they leave elementary school and progress through high school and into postsecondary institutions or the work force. The base year study, conducted in the spring term of the 1987-88 school year, selected 26,432 potential eighth grade participants, of whom 24,599 were successfully surveyed in 1,052 public and private schools. Additional data were gathered from eighth graders' parents, teachers, and principals.

The *first follow-up* in 1990 provided the first opportunity for longitudinal measurement of the 1988 baseline sample. It also--after sample freshening[1]--provided a comparison point to high school sophomores ten years before, as studied in High School and Beyond (HS&B). One of the chief goals of the NELS:88 design has been to capture in a longitudinal data set the entire subpopulation of school dropouts from within a high school entry cohort.

The *second follow-up* took place in 1992, when most sample members entered the second term of their senior year. The second follow-up provides a culminating measurement of learning in the course of secondary school, and also collects information that will facilitate investigation of the transition into the labor force and postsecondary education after high school. Freshening the NELS:88 sample to represent the twelfth grade class of 1992 makes trend comparisons with the senior cohorts that were studied in the National Longitudinal Study of the High School Class of 1972 (NLS-72) and HS&B possible. The NELS:88 second follow-up resurveyed students who were identified as dropouts in 1990, and identified and surveyed those additional students who left school after the first follow-up.

In all three in-school rounds, students were asked to complete a student questionnaire, taking up to one hour to do so, and an 85-minute achievement battery comprising tests in reading, mathematics, social studies and science. Except for some experimental constructed response items in 1992 (on which, see Pollack and Rock, 1996), the cognitive tests were in a multiple choice format.

The NELS:88 *third follow-up* took place in the spring of 1994, with most sample members in postsecondary education or in the labor force. Base year ineligible students who remained ineligible in 1990 and 1992 were not included in the 1994 round. Base year ineligible students who had become eligible in 1990 or in 1992 are represented in the 1994 sample. A major goal of the 1994 round is to provide data for trend comparisons with NLS-72 and HS&B, and to continue cross-wave comparisons with previous NELS:88 rounds. The 1994 survey provides a basis for assessing how many dropouts have returned to school and by what route, and measures the access of dropouts to vocational training programs and to other postsecondary institutions. A *fourth follow-up* is scheduled for 1998 or 2000.

Sample Undercoverage in NELS:88. Errors in surveys are generally classified into two types: sampling and nonsampling errors. Sampling errors consist of differences between the sample and the population that are attributable to chance. Nonsampling errors in surveys and censuses may be grouped in three broad categories: (1) measurement or response errors, (2) errors due to nonresponse or missing

[1] The process of "freshening" added spring term 1990 sophomores who were not in the base year sampling frame, either because they were not in the country or because they were not in eighth grade in the spring term of 1988. The 1990 freshening process that provided a representative NELS:88 sophomore cohort was repeated in 1992 in order to provide a NELS:88 senior cohort, that is, a nationally representative sample of students enrolled in twelfth grade in the spring term of 1992.

data, and (3) coverage errors. This report is concerned with coverage error. Simply stated, coverage error is the failure to include the entire universe of interest in the population to which the sample data are projected. Coverage error is normally caused by an incomplete or out-of-date sampling frame (for example, school lists or student rosters that are inaccurate or have not been updated), but it may also arise from eligibility rules that are inappropriate or not reliably and validly applied. In addition, the decision to exclude a portion of the universe as not of interest may always be queried and must always be justified. The issue of coverage ultimately leads back to the question of how the universe of interest should be defined.

There are many motivations for excluding categories of students *by design* from assessments and surveys. Sometimes students are excluded because of the practical and cost difficulties of including them, even though, conceptually, they are part of the population of interest. In other instances exclusion is based on the belief that the assessment (or questionnaire) would not be valid for a student within some given classification, either because the student's disability or limited proficiency in English would not permit accurate measurement using the assessment instruments, or because the student's classification is tied to a separate curriculum with distinct curricular goals such that the content of the assessment would be inappropriate. If a student receives a different curriculum from the one an achievement test measures, then the argument for regarding that student as legitimately not part of the defined "population of interest" would seem to be compelling. Indeed, some students with disabilities -- the severely impaired --- are in a separate curriculum. Most are not, nor, by and large, are LEP students although the unique features of bilingual-bicultural education programs should be taken into account. There is also an issue of *who* interprets exclusion criteria once they are set. Since assessments and surveys depend on school personnel to apply eligibility rules, the school-level incentives and disincentives to include or exclude particular categories of students must also be considered, as well as the schemes of classification and culture of the school that will influence the judgments of those implementing exclusion guidelines.

This methodology monograph addresses issues of student sample exclusion in NELS:88, particularly the problem posed by eighth graders who were declared ineligible for the study owing to factors -- physical or mental disabilities, or lack of proficiency in the English language -- deemed by their schools to pose a significant barrier to participation. In the base year, 5.4 percent of eighth graders were so excluded. (The *weighted* proportion of the sample that was excluded is 4.7 percent.) A sample of the excluded students was followed in subsequent rounds, and eligibility status reassessed. A substantial number of excluded 1988 eighth graders re-entered the NELS:88 sampling frame as eligible in 1990 and 1992.

Organization of This Report. This report is divided into five chapters, a bibliography, and five appendices:

Chapter 1:

- provides background on base year student sampling;
- discusses the exclusion of students as a potential source of undercoverage;
- discusses the reliability and validity of exclusion criteria;
- describes the objectives and the implementation of the first follow-up Base Year Ineligibles study, and the second follow-up Followback Study of Excluded Students;

Chapter 2:

- describes the characteristics of the base year ineligibles, specifically, the proportions belonging to various racial/ethnic or gender groups;
- describes the proportion having selected school outcomes such as persistence in school, remaining in or falling behind modal grade progression;

Chapter 3:

- depicts changes in ineligibility status between 1988 and 1992;

Chapter 4:

- discusses other possible sources of undercoverage in the NELS:88 sample;

Chapter 5:

- outlines conclusions drawn from this study and recommendations for future studies;

the Bibliography:

- details references used in preparing the report;

Appendix A:

- provides supplemental tables that include data for all figures and tables reported in chapters 1-5 and additional bivariate statistics not reported in the main text;

Appendix B:

- contains examples of screeners used in the redetermination of eligibility;

Appendix C:

- provides information about eligibility criteria employed in HS&B and NAEP;

Appendix D:

- provides technical notes on the sample, approach to analysis and survey methodology;

Appendix E:

- supplies a glossary of technical terms used in this report.

Summary

Who Was Excluded?

- About five percent of eighth graders in the base year were excluded from participation by their schools
- Of the excluded 1988 eighth graders, 20 percent were ineligible because their English language proficiency was deemed to be insufficient for survey participation; 66 percent were classified as ineligible owing to mental disabilities; 6 percent were excluded due to physical disability; and 8 percent were classified as "disability unknown".
- Some 63.5 percent of the excluded eighth grade cohort were males, and 36.5 percent females
- Of 1988 eighth graders excluded owing to a mental disability, 30.7 percent were female and 69.3 percent male; among those excluded because of limited English language proficiency, 49 percent were female versus 51 percent male
- Of the eighth grade population excluded because of a mental disability, about 59.8 percent were white, 19.4 percent black, 8.3 percent Hispanic, and 0.6 percent Asian; the remainder were of unknown or other race
- Only 62.4 percent of the eighth grade cohort 1988-ineligible population were still in high school four years later and 37.6 percent were not
- Of the 37.6 percent not in high school, however, some 7.6 percent were taking some sort of educational program in preparation for an alternative educational credential such as the GED, while the remaining 30 percent were dropouts receiving neither regular nor alternative instruction
- Of the 62.4 percent of the excluded 1988 eighth grade cohort population that was in school in 1992, 57.6 percent were in modal grade sequence, that is, were spring 1992 high school seniors and 42.4 percent were not

How Many Could be Included, Four Years Later?

- ,57 percent were now eligible, 32 percent remained ineligible, and for 11 percent, status could not be ascertained owing to locating problems
- 71 percent of the LEP excluded students had become eligible
- 50 percent of IEP students with a mental impairment had become eligible

How Might Inclusion be Meaningfully Maximized?

This report contains recommendations in thirteen areas for ways in which more complete and accurate information could be collected for all categories of students. These areas are:

- Provision of modifications or special accommodations in testing and questionnaire administration
- Establishment of an appropriate comprehension and measurement floor for student assessments and questionnaires
- Use of parallel verbal/nonverbal items in assessment
- Use of performance assessments and teacher ratings
- Test and questionnaire translation into other languages
- Use of ancillary data (school records, contextual data questionnaires from principals, teachers, parents)
- Imputation of missing test data
- Reassessment of eligibility status over time
- Sensitivity of inclusion strategies to age of the cohort
- IEP/LEP oversampling or parallel special studies
- Refinement of eligibility definitions and protocols
- Establishment of consistent markers of LEP and IEP status and of disability type and consistency in eligibility rules across databases, in order to facilitate cross-study and trend comparisons
- Conduct special substudies of excluded students

Chapter 1: Excluded Students[2] In the NELS:88 Base Year

1.1 Background

In the base year of NELS:88, students were sampled through a two-stage process. First, stratified random sampling and school contacting resulted in the identification of the school sample; second, students were randomly selected (with oversampling of Hispanics and Asians) from within cooperating schools.

The target population for the base year comprised all public and private schools containing eighth grades in the fifty states and the District of Columbia. Excluded from the NELS:88 school sample were Bureau of Indian Affairs (BIA) schools, special education schools for the handicapped, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas.[3] The sample excluded those students who, in the judgment of school personnel, faced significant barriers to participation, specifically, students with severe mental disabilities, insufficient command of the English language to understand the survey materials (especially the cognitive tests), or physical or emotional concerns that would render participation unduly difficult.

1.1.1. Exclusion of students. The target student population for NELS:88, about which we would ideally like to make statistical inferences, consists of all eighth graders enrolled in public or private schools in the fifty states or the District of Columbia in the spring term of the 1987-88 school year. To better understand how excluding students with disabilities or language barriers affects population inferences, data were obtained on the numbers of students excluded as a result of these restrictions.

Seven ineligibility codes defining categories of excluded students were employed at the time of student sample selection:

A - attended sampled school only on a part-time basis, primary enrollment at another school.

B - physical disability precluded student from filling out questionnaires and taking tests.

C - mental disability precluded student from filling out questionnaires and taking tests.

D - dropout: absent or truant for 20 consecutive days, and was not expected to return to school.

E - did not have English as the mother tongue AND had insufficient command of English to complete the NELS:88 questionnaires and tests.

F - transferred out of the school since roster was compiled.

G - was deceased.

---

[2] The terms (1988) "excluded student" and "base year ineligible" (BYI) are used interchangeably in this report.

[3] For further details of school-level exclusion, see Spencer, Frankel, Ingels, Rasinski and Tourangeau, 1990, p. 10.

Before sampling, school coordinators--members of the school staff, typically an assistant principal or guidance counselor who acted as liaison between the school and the study--were asked to examine the school sampling roster and annotate each excluded student's entry by assigning one of the exclusion codes.[4] Because eligibility decisions were to be made on an individual basis, special education (IEP)[5] and Limited English Proficiency (LEP)[6] students were not to be excluded categorically. Rather, each student's case was to be reviewed to determine the extent of limitation in relation to the prospect for meaningful survey participation. Each individual student, including LEPs and students with physical or mental disabilities, was to be designated eligible for the survey if school staff deemed the student capable of completing the NELS:88 instruments, and excluded if school staff judged the student to be incapable of doing so. School coordinators were told that when there was doubt, they should consider the student capable of participation in the survey. Exclusion of students after sampling ("post-roster ineligibles") occurred either during the sample update just prior to survey day, or on survey day. Exclusion after sampling normally occurred because of a change in student status (for example, transfer, death). However, in very rare instances such exclusions reflected belated recognition of a student's pre-existing ineligibility--that is, if an annotation error was made and an ineligible student selected for the sample in consequence of such an error, ineligibility became apparent later in the survey, whereupon the student was excluded.

---

[4] In some schools, some students were assigned multiple ineligibility codes. In these rare cases, school coordinators were instructed to assign one of the ineligibility reasons as primary.

[5] Special education students are individuals with mental or physical disabilities who, normally, have on file an Individualized Education Plan (IEP). Presently over five million elementary and secondary school students (nearly 12 percent of the public school population) have IEPs that make them eligible for special education services.

[6] Limited English Proficient students are contrasted with No English Proficiency (NEP) students and those who are Fully English Proficient (FEP). For purposes of this report, LEP and NEP students are lumped together as a single category. Limited English Proficiency is a categorization developed to assist in identifying those language-minority students (individuals from non-English language backgrounds) who need language assistance services, in their own language or in English, in the schools. The Bilingual Education Act, reauthorized in 1988 (PL 100-297), describes a limited English proficient student as one who:

1) meets one or more of the following conditions:

   a) the student was born outside of the United States or the student's native language is not English;

   b) the student comes from an environment where a language other than English is dominant; or

   c) the student is American Indian or Alaskan Native and comes from an environment where a language other than English has had a significant impact on his/her level of English language proficiency; and

2) has sufficient difficulty speaking, reading, writing, or understanding the English language to deny him or her the opportunity to learn successfully in English-only classrooms.

Owing at least in part to local definitional differences, estimates of the number of LEP students differ. Some sources indicate a population of 3.5 to 5 million school-age LEP students (CCSSO, 1990). Recent tabulations (Henderson, Abbott & Strang, 1993) put the number of LEP students at 2.371 million in kindergarten through twelfth grade for the year 1991-92 in the fifty states and the District of Columbia, or at 5.6 percent of enrollment (6.1 percent of public school and 1.1 percent of private school enrollment). At eighth grade level, NAEP in 1992 identified three percent of students as LEP and excluded two percent as LEP; at grade four, four percent were identified as LEP and three percent excluded (Mullis, Dossey, Owen & Phillips, 1993). It should be noted, however, that there are also native speakers of English, who may or may not have learning disabilities, whose language proficiency is very limited -- this may be especially the case for reading ability.

Excluded students were divided into those who were full-time students at the school (categories B, C, and E) and those who were not (categories A, D, F, & G). Our main concern here is with students who were full-time students at the school but who were excluded from the sample. Excluding these students affects estimates made from the sample.[7]

The total eighth grade enrollment for the NELS:88 sample of schools was 202,996. Of these students, 10,853 were excluded owing to limitations in their language proficiency or to mental or physical disabilities. Thus 5.37 percent of the potential student sample (the students enrolled in the eighth grade in the 1,052 NELS:88 base year schools) were excluded. In the 1988 National Assessment of Educational Progress (NAEP), about the same proportion of eighth graders (for NAEP, 5.3 percent) were excluded.[8] Less than one half of one percent of the potential sample was excluded for reasons of physical or emotional disability (.41 percent), but 3.04 percent was excluded for reasons of mental disability, and 1.90 percent because of limitations in English proficiency.[9] Put another way, of the 10,853 excluded students, about 57 percent were excluded for mental disability, about 35 percent owing to limitations in English language proficiency, and less than 8 percent because of physical or emotional disabilities.[10]

1.2 Exclusion as an Undercoverage Problem

In the NELS:88 documentation generally, we have referred to the exclusion of a sizable proportion of students with disabilities and students with limited English proficiency as an undercoverage problem. In a technical sense, one can argue that this is not so -- these students were defined as outside the target population, and talk of coverage error might more appropriately be limited to missed or duplicate units on the sampling frame that lead to discrepancies between statistics calculated on the frame population and on the target population (see Groves, 1989, p.83). Nonetheless, we have viewed exclusion as an undercoverage problem because we believe that, given the purposes of NELS:88, the base year eligibility rules seriously misspecified the target population. While it was in principle appropriate to exclude some students from the cognitive testing component of the study, most students who could not be tested could nevertheless have participated in NELS:88. These individuals could have completed a self-administered questionnaire or could have been interviewed by field personnel, and their parents, teachers, and principals could have been surveyed, and their school records obtained. (The NELS:88 design called for obtaining and analyzing a great number of outcome measures, not just achievement test scores.) Compounding this misspecification of the target population is the fact that the eligibility rules, whatever

[7] Students in categories A (n=329), D (n=733), F (n=3,325), and G (n=6) were either not at the school or were present only part time (with primary registration at another school, hence a chance of selection into NELS:88 at another school). Excluding students in these categories has no implications for making estimates to the population of eighth grade students. Students in category F, those who had transferred out of the sampled school, had some chance of being selected into the sample if they transferred into another NELS:88 sampled school just as transfers into NELS:88 schools from non-NELS:88 schools had a chance of selection at the time of the sample update. The sampling of transfer-in students associated with the sample update allowed NORC to represent transfer students in the NELS:88 sample.

[8] While 5.3 percent of eighth graders were excluded in the 1988 assessment, NAEP excluded 3.7 percent of twelfth graders and 6.3 percent of fourth graders. Exclusion rates are typically higher for younger age/grade groups.

[9] These are (unweighted) *sample* proportions. Proportions that are weighted *population estimates* differ. Also, some cases included in the unweighted proportions (N =48) were subsequently discovered to be sampling errors, shrinking the ineligibles sample by about 7 percent.

[10] Again, weighted proportions differ; this is most markedly true for proportion of language excluded, and reflects the oversampling of Asians and Hispanics.

their merits, appear to have been inconsistently applied, leading to the inappropriate exclusion of students from the survey, including exclusion of students who could have been tested. It is for these reasons that we speak, even if somewhat unconventionally, of coverage error in connection with the exclusion of categories of students from NELS:88.

There are a number of statistical concerns associated with excluding a portion of the student population. In particular:

- Exclusion biases overall estimates;
- Exclusion severely limits the ability to generalize about and compare to the general population a number of subpopulations of intense policy interest such as students with limited English proficiency or students with disabilities;
- Because eligibility status changes over time, baseline exclusions, unless re-addressed in future rounds, threaten the representativeness of subsequent freshened grade-representative cohorts within a longitudinal design.

It may be desirable to elaborate on these points. First, because current characteristics and probable future educational outcomes for disproportionately excluded groups may depart from the national norm, the exclusion factor should be taken into consideration in generalizing from the NELS:88 sample to eighth graders in the nation as a whole. This implication for estimation carries to future waves. For example, if the overall propensity to drop out between the eighth and tenth grades is higher for excluded students than for base year-eligible students, the dropout figures derivable from the NELS:88 first follow-up (1990) will underestimate early dropouts. This in fact is this case; the table below shows the difference between estimates of the national dropout rate for the 1988 eighth grade cohort when estimation is grounded in the eligible-only sample, as contrasted to all members of the cohort (as represented by the "expanded sample").

Table A: Bias Estimates for Eighth Grade Cohort Dropout Rate, 1988-1990.
(Percentage of Spring Term 1988 Eighth Graders Not in School Spring Term 1990)

	ELIGIBLE SAMPLE		EXPANDED SAMPLE		BIAS
Total.	6.0 %	(0.48)	6.8 %	(0.40)	**-.8**
Race/Ethnicity.					
Asian	3.1	(1.05)	4.0	(1.02)	**-.9**
Hispanic	9.2	(1.01)	9.6	(0.84)	**-.4**
Black	10.0	(1.94)	10.2	(1.51)	**-.2**
White	4.9	(0.53)	5.2	(0.44)	**-.3**
Sex.					
Male	6.3	(0.69)	7.2	(0.55)	**-.9**
Female	5.8	(0.59)	6.5	(0.51)	**-.7**
1988 Eighth Grade Public School Students.	6.8	(0.55)	7.6	(0.45)	**-.8**

Note: **Standard errors appear parenthetically after each estimate. Two small subgroups do not appear under race/ethnicity. One such group, race unknown, comprised about 2 percent of the unweighted expanded sample, and had by far the highest dropout rate of any group. The second group, American Indians, comprised over 1 percent of the sample.**

Source: **National Education Longitudinal Study of 1988 (NELS:88) First Follow-Up, National Center for Education Statistics, public use file and expanded cohort file.**

Second, exclusion undermines the power to characterize the subgroup of students with disabilities or limited English proficiency. One can, however, generalize about students whose lack of proficiency or whose disability is less severe.

Third, in a school-based longitudinal survey such as NELS:88, baseline excluded students potentially affect future representativeness of freshened grade cohorts. To achieve a thoroughly representative sophomore (1990) and senior (1992) sample comparable to the High School and Beyond 1980 sophomore cohort (or, for 1992, the HS&B 1980 senior cohort and the base year of NLS-72), the NELS:88 follow-up samples must approximate those which would have come into being had a new baseline sample independently been drawn at either of the later time points. In 1990 (and 1992) one must therefore freshen, to give "out of sequence" students (for example, in 1990, those tenth graders who were not in eighth grade in the spring of 1988) a chance of selection into the study. One must also accommodate excluded students whose eligibility status has changed, for they too (with the exception of those who fell out of sequence in the progression through grades) would potentially have been selected had a sample been independently drawn two years later, and must have a chance of selection if the

representativeness and cross-cohort comparability of the follow-up sample is to be maintained. Thus, for example, if a base year student excluded because of a language barrier achieves the level of proficiency in English that is required for completing the NELS:88 instruments in 1990 or 1992, that student should have some chance of re-entering the sample.

A fourth concern is that these undercoverage problems may arise even if eligibility criteria are conceptually appropriate and well-defined. A further problem arises if eligibility rules are vague or inappropriate or are difficult to apply in consistent fashion, if those who apply the rules do not have sufficient information, or if they have an incentive to exclude students on any basis other than an impartial reading of available facts.

1.3 Reliability and Validity Issues: Were the Base Year Exclusion Criteria Applied Consistently?; Are Excluded Students Unable to Participate?

The objectives of the followbacks of NELS:88 excluded students are described in section 1.4 below, but it is important to note at this juncture that one possible objective that was not pursued was a rigorous scientific test (for example, an experiment or quasi-experiment) of the reliability, or the validity, of exclusion procedures. Nevertheless, the experience of the NELS:88 base year and the results of the first and second follow-up studies of excluded students support the notion that eligibility criteria are not always applied from school to school in a highly consistent manner.

School incentives to exclude students. There have been a number of recent discussions of the extent, reliability, validity, and implications of student exclusion from major national data bases.[11] These accounts suggest that greater inclusiveness in assessments and surveys might be achieved.[12] Schools have a number of incentives to exclude students from participation in data collection programs, and it is important to understand these incentives in order to know how to maximize meaningful participation.

One incentive to exclude students from testing programs is that LEP and IEP students will, on average, pull test scores down. Low-achieving schools within an accountability system that relies on assessment results have an incentive to remove children from the assessment stream by classifying them as having mental disabilities and as students who should not be tested (McGill-Franze & Allington, 1993). While tests such as the NELS:88 achievement battery are not part of a district or state accountability system, schools will generally avoid putting students in testing situations who already have been classified in such a way that they are not formally to be assessed. This same incentive to exclude applies with equal force to LEP students.

---

[11] See, in particular, McGrew, Thurlow, & Spiegel, 1993; Ysseldyke & Thurlow, eds., 1993; and Spencer, 1991.

[12] McGrew, Thurlow, and Spiegel, 1993, p. 345 estimate that 36 percent of students with disabilities were excluded from the NELS:88 sample. Among students with disabilities, however, while 10 percent nationally are classified as severely handicapped, 90 percent are classified as mildly handicapped (see, for example, Algozzine and Korinek, 1985; and D.J. Reschley in Ysseldyke and Thurlow, eds., 1993.) Reschley notes that "the vast majority of students with disabilities do not have identifiable biological anomalies that would interfere with participating in state and national assessment programs" and suggests that probably "less than two percent of the overall student population has a biological anomaly that would interfere with performance on assessment procedures like group administered standardized tests." (p.38). Such evidence suggests the possibility that more excluded students could be meaningfully included. At the same time, such evidence also suggests that perhaps two percent of students owing to disabilities and perhaps an additional one percent of high school students (and more, in early grades) owing to language barriers, cannot, under normal survey conditions, be readily included in testing programs.

Availability of categorical funding may influence student classification, hence, ultimately, the pool of students schools feel should be tested. The size of the special education population is currently growing at a faster rate than the increase in school-age population. Possible reasons for this rate of increase include the greater proportion of children living in poverty, increase in prenatal exposure to drugs, and "fiscal incentives for identifying students in need of supplemental services" (OSEP 1994, p.7). Some types of disabilities, such as sensory or motor deficits, may (generally) be easily identified. Placements for disability categories with softer diagnostic criteria (such as the most common disability, learning disability) are more likely to be affected by the availability of different types of funds for special services. Most of the current disproportionate growth in identification of special education students is in the learning disability classification (OSEP 1994, p.25). Schools find it difficult in any case to distinguish between children with learning disabilities and those who are low achievers (Clay, 1990; Kronick, 1988; Reynolds, 1990). Availability of funding to assist the education of learning disabled students provides an incentive for maximum identification -- or over-identification of slow learners as learning disabled.[13] IEP teams, in turn, have an incentive to exclude students with disabilities from general assessments, from which the student receives no direct benefit, and completion of which the student may find taxing and difficult. (For that matter, few low-achievement students, regardless of whether they are classified as having a disability, would be likely to find completion of a test such as the 85-minute NELS:88 cognitive battery anything less than onerous)

Another factor that may lead to students being excluded who could in fact complete assessments or questionnaires is reliance on general categories, such as LEP and IEP that are themselves subject to error or indeterminacy, but that are used by schools as a shortcut or alternative to making an individual-level decision. Above, we noted the difficulty that schools have in distinguishing low achievers and individuals with learning disabilities. Bennett and Ragosta (in Willingham et al., 1988, p.20) describe the special education and classification process as follows:

> Typically, students enter the process as a result of referral by their classroom teachers. Research has shown that such referrals are sometimes based on such extraneous factors as race, sex, physical appearance, and socioeconomic status....Assessment is followed by a classification meeting at which a diagnosis is made. Investigations of this aspect of the process report little consistency in diagnostic statements among professionals assigned to the same case, only a slight relationship between assessment data and team judgments, and the influence of irrelevant pupil characteristics on classification decisions....As a result, studies suggest that over half of the classification decisions made by child-study teams are erroneous (Algozzine & Ysseldyke, 1982; Craig, Myers, & Wujek, 1982; Shepard, Smith, & Vojir, 1983). One effect of these placement errors is to confuse attempts to characterize the true nature of handicapping conditions even further.

Further evidence of the inconsistent application of disability categories is the state-to-state variation (1992-93 school year) in students identified for special education. Nationally, .89 percent of students are classified as having serious emotional disturbance, but the proportion in Connecticut is 2.1 percent and

[13] Rates of disability identification are demonstrably sensitive to funding formulas. The *Sixteenth Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act* (OSEP 1994, p.26) notes: "OSEP is currently funding a small study to examine reasons for the growth in the number and percentage of students identified with specific learning disabilities. Many State and local agencies are also taking steps to control growth in the special education population, primarily through revisions in State funding formulas. For example, changes to the funding formula in Vermont have curbed the growth of the special education population."

in Mississippi .04 percent. Nationally, about 5.25 percent of students are classified as learning disabled, but states range from 9.4 percent (Massachusetts) to 2.7 percent (Georgia). Large differences are seen even for disabilities that have "hard" objective criteria; for example, nationally .13 percent of students are classified as having hearing impairments but .26 percent of students in Oregon are so classified and only .02 percent in Florida.

Nor is there a consistent, determinate definition of LEP that is everywhere followed.[14] A usable definition of LEP for assessment and survey classification purposes, moreover, would reflect the continuum of skill levels across reading, writing, listening, and speaking. *The Annual Evaluation Report: Fiscal Year 1990* of the U.S. Department of Education's Office of Planning, Budget and Evaluation describes research on LEP placement:

> A study of student selection procedures found that, when a student speaks some English, different oral language proficiency tests often disagree as to whether the student should be classified as LEP. Classification of such students as LEP depends on what test is used and how high or low a local district or state chooses to set cut-off scores for selection into or exit from the program. (See Pelavin, 1988).
>
> The recently completed National Longitudinal Evaluation of Bilingual Education found the major determinant of LEP placement in particular instructional services to be the school district's policies, some of which are apparently independent of individual students' level of English proficiency or teacher judgments. These policies are related to local conditions, such as the size of the LEP population in the district, school and classroom. English proficiency of the LEP student does play a minor role in assignment to service, but mostly in the earlier grade levels. Decisions about when to exit students from LEP services are based on both local policy determinations and rules related to reaching certain levels of English proficiency. (See Burkheimer et al., 1990)

Apart from the fact that the meaning of "LEP" may vary from place to place, it may typically be easier for a school administrator to assume that students in, for example, a bilingual education program, cannot be assessed in English. The alternative is a far more laborious process of individually assessing each program participant's English language literacy, either directly or through solicitation of the expert advice of individual teachers.

To be sure, in any case of the application of general criteria, there is bound to be some degree of arbitrariness in judgments about borderline cases. This arbitrariness is of course compounded when the numbers of people (over a thousand individuals in the NELS:88 base year) rendering eligibility judgments is large, when the number of individuals to be judged is also large (over 200,000 eighth graders), and when the individuals making the judgments are typically not personally familiar with the capacities of the individuals whose ability to complete the survey forms is to be judged. Our greatest concern about the classification process is that, for reasons of time and burden, some schools apparently departed from their instructions and excluded students on a categorical basis in preference to rendering the prescribed case-by-

[14] On the lack of consensus on the concept and definition of language proficiency or competence, see Gandara and Merino, 1993.

case assessments.[15] In consequence both of inconsistencies in application of eligibility criteria and of categorical exclusion, one would expect that overall, more students may have been excluded than necessary. The temptation to exclude categorically--in a school with a large eighth grade, given severe time pressures for producing an annotated roster, and with individual-level information available to the School Coordinator only through the laborious process of interviewing the special education or English as a second language or bilingual education teacher of each student--is large. The methodology (described in detail below) was designed to minimize this problem in the excluded student followbacks. We sought greater precision in exclusionary definitions, and sought guidance from special education, English as a second language, and bilingual education teachers.

1.4 Studies of the Base Year Ineligibles in NELS:88 First and Second Follow-Ups: Objectives

The longitudinal follow-up of excluded 1988 eighth graders undertaken in the NELS:88 first follow-up (and repeated in the 1992 second follow-up[16]) was designed to realize several important aims.

- First, to increase the generalizability of key population inferences, particularly, adjusted 1988-90 and 1988-92 dropout rates that reflect full coverage of the eighth grade cohort.
- Second, to depict more fully the characteristics of students excluded from sample estimates so that more explicit caveats can be entered about the limitations of sample coverage in NELS:88.
- Third, to enhance the representativeness of the high school-based NELS:88 follow-up samples by accommodating change in the eligibility status of the eighth grade cohort over time.
- Fourth, to correct errors in the application of base year eligibility criteria by (a) providing more specific guidelines and (b) seeking the input of individuals such as teachers with specific knowledge of the capabilities of the excluded student.
- Fifth, to enforce consistency in eligibility definitions between the base year and follow-ups of NELS:88, by applying the broadened eligibility criteria of the First Follow-Up (which extended eligibility to students who could complete the questionnaire only in Spanish) to excluded 1987-88 eighth graders.

In the second follow-up, 1990 sophomores who were declared ineligible for the sample in the freshening process were examined as well as 1988 ineligible eighth graders, since the eligibility status of these excluded students, too, might have changed (for example, a 1990 sophomore with language limitations might have gained substantially in English proficiency by 1992).

---

[15] Evidence for this phenomenon was seen when sampling rosters were inspected at the beginning of the first follow-up Base Year Ineligibles Study, and rosters were found on which all students within a pre-existing category were excluded. Further evidence for this had been uncovered during the base year. When rosters were returned that had an extraordinarily high number of exclusions, we typically called back the school to find out why. In most cases, exclusion was being applied categorically. In such cases, we attempted (often but not always successfully) to persuade the school to assume the extra burden of individualized classification and re-annotate the rosters.

[16] In the first follow-up (1990), this group was investigated in the Base Year Ineligibles Study. In the second follow-up (1992), this group was investigated in the Followback Study of Excluded Students.

While demographic and status information was gathered for all members of the excluded student followback sample, three situations justified inducting a formerly excluded student into NELS:88 and administering the student questionnaire (and tests, whenever possible, in 1992). The three situations are as follows: that person (1) had changed, that is, now met the eligibility criteria; (2) was wrongly classified in 1988; (3) was rightly classified and the student's limitation was unchanged, but this person met broadened eligibility criteria (that is, could complete the student questionnaire in Spanish).

1.5 The Base Year Ineligibles Studies: Implementation[17]

1.5.1. Sampling. The sample of ineligible students was drawn from the 1,052 fully participating Base Year schools. Students were included who had been excluded by virtue of a physical or mental disability, because of a language barrier to participation, or who had been excluded with no reason given. Excluded students were sampled using the school's original selection table, following the next unused number(s) on the table. This procedure resulted in a sample frame of 1,598 students. NORC selected a subsample of 674 to be studied in the first follow-up.

1.5.2. Instrumentation. For all base year ineligibles in the excluded student followback sample, the following status information was to be obtained from the student's current school (if enrolled) or school last attended (if a dropout) upon screening:

- Sex: male or female;
- Race/ethnicity: white, black, Hispanic, Asian, American Indian, other
- School enrollment status

Students were next screened for eligibility. (This process is described below; the eligibility screener is reproduced in Appendix B).

For students classified as still ineligible, no further information was collected, beyond locating data to facilitate future follow-up and a detailed description of the precise reason for continued (1990) ineligibility. For students deemed to be eligible, the first follow-up student questionnaire and new student supplement were administered. For cost reasons, test administrations were deferred until the second follow-up.

Ineligible 1988 eighth graders who were again classified as ineligible in 1990, or sophomores who were selected for the freshening sample but deemed ineligible, were included in the second follow-up Followback Survey of Excluded Students, which employed a like methodology and reassessed eligibility status for the spring term of the 1992 school year.

1.5.3. Eligibility Screening. In the base year, school personnel--typically the school principal or the school-appointed coordinator for NELS:88--reviewed rosters and indicated students who should be excluded owing to mental, physical, or language barriers to completing the NELS:88 survey forms. For the base year ineligibles followback, NORC attempted to gain information from a teacher or counselor

[17] For the first follow-up, a more detailed account of sampling and data collection methodology can be found in the *NELS:88 First Follow-Up Final Technical Report* (Ingels, Scott, Rock, Pollack, & Rasinski, 1994). For the second follow-up, see Ingels, Dowd, Baldridge, Stipe, Bartot & Frankel, 1994, sec.4.3.6.

who had extensive personal knowledge of the student and the student's school situation, and we attempted to provide more specific guidelines to help school personnel to weigh whether a given individual was capable of participation.

In determining eligibility status as of spring term 1990, interviewers were instructed to obtain reports from a person with first-hand knowledge of the student. It was not sufficient simply to talk to someone in the school office, or the principal. Interviewers were to approach the special education teacher, the bilingual education or language arts teacher, or other relevant individuals who had first-hand knowledge of the excluded student's academic capacities. This process typically entailed talking to multiple staff members of the school, until the individual best qualified to assess the student's eligibility status was identified.

Special education personnel are often highly protective of individuals with IEPs, and sometimes place more emphasis on whether completing the survey forms will benefit the individual than on how the individual's participation would benefit the research program at hand. We therefore stressed to special education instructors the right of all students who are capable to participate, and the importance of making national data representative of all populations served by the nation's schools. At the same time, we stressed as well that students for whom participation truly would be unduly burdensome, either physically or psychologically, or not meaningful, must be excluded.

Eligibility criteria appear on the first follow-up eligibility screener in Appendix B. Some of the criteria conform closely with the eligibility rules adopted for NAEP[18] in 1990, while others depart from the NAEP guidelines. The language inclusion and exclusion guideline follows the NAEP model (normally, sample members who have been enrolled in an English-language course of study for at least two years would be considered eligible), though with an additional provision for participation in the form of a Spanish-language questionnaire. The inclusion and exclusion guideline for students with disabilities attempts to define an objective ability floor for reading comprehension, to encourage special education teachers to include any student who could read at a level above the sixth grade norm.[19] In addition, we placed redoubled emphasis on the injunction that when school personnel were in doubt, they should include, and to further reduce incentives to exclude, stressed that the NELS:88 tests were *tests for national statistics*, not *tests for accountability*. The school's score would not be reported as such and the school's identity in the public data files could not be deduced, hence the school's results could not be reported or judged within a state or local accountability system. Nor would results be used in a "high stakes" way for students. Student results would be employed descriptively in statistical summary and research and would be cloaked in anonymity.

1.5.4. Modifications or Accommodations to Include IEP and LEP Students. In general, our approach assumed that special testing accommodations would not be made, although rare exceptions could be allowed when it was possible to be reasonably sure that such adaptations would not diminish the comparability of results. In terms of other survey forms such as questionnaires, however, a special

[18] See, for example, Mullis, I., 1990, *The NAEP Guide: A Description of the Content and Methods of the 1990-92 Assessments*. Washington, D.C.: NCES. Also see Appendix C of this report.

[19] The issue of "how low is the floor?" is relevant primarily to learning-disabled students, as well, perhaps, to poor readers generally; many physically handicapped students, for example, are outstanding achievers. The NELS:88 eighth grade tests contained some third grade items and many grade 4 - 6 items; the follow-up tests tended to raise the ability ceiling but to do little to change the floor. Tests were not administered to the reclassified ineligibles in 1990, but were to be administered in 1992. The questionnaire was also designed to be understandable to students who read several levels below their grade norm.

BEST COPY AVAILABLE

accommodation that was made for LEP students was, in 1990 and in 1992, a Spanish-language translation of the student questionnaire.

Accommodations. There are a number of special accommodations that can be made that would extend the number of individuals who could meaningfully complete survey questionnaires and cognitive tests. Special accommodations to facilitate participation in direct assessments (all of these adjustments are also effective for facilitating questionnaire administration) include the following: extended time limits or breaking the test into multiple sessions; small group or one-on-one administration; paraphrase, reading to the student (directions, or questions and content), using visual aids; allowing the use of dictionaries; taking dictation from the student; providing special acoustics, furniture, visual magnification or auditory amplification devices; and providing large print versions (or magnifying glasses) or Braille versions of instruments (or Braille or "talking" calculators). An additional option--translation--is discussed separately, below.

In 1992, when base year excluded students were subject to being tested, the only accommodations made for IEP students were (very rarely) one-on-one administration for emotionally disturbed students who might be distracted by or disruptive in a group administration setting, and in one instance, enlargement of the survey forms so that a larger print version could be utilized. In part this assumption that special accommodations would not normally be made reflected cost considerations, but in larger measure, validity considerations as well -- it was not clear that modified administrations of the NELS:88 tests would produce truly equivalent results. Little work has been done on these important validity questions at this time[20] and nothing specific to the NELS:88 tests or close analogues to them.

Translation. Translation, including the use of sign language for individuals with auditory disabilities, is an important option. No attempt was made to translate the NELS:88 achievement battery into another language. Apart from the fact that LEPs speak many dozens of languages (for example, in New York City, although 68 percent of LEPs served by the public schools speak Spanish, the city's LEPs have more than 130 different mother tongues), there are many problems associated with the translation of tests, and it is difficult to achieve assurance of their equivalence.[21] (However, scores could be reported separately and still be of utility for some purposes.) It is important that further work in translation be done. In addition to the contribution that might be made to assessment within the United States, international education comparisons also hinge on the issue of achieving equivalent forms in multiple languages. Nevertheless, translations of achievement tests make sense, at best, only if they are in the student's primary language of instruction. The majority of native speakers of Spanish in the United States are instructed in English. If one wants to measure achievement in the context of a classroom

[20] There are several classes of accommodation. Some involve alternative presentation or response modes (e.g., braille, sign language, large print, oral reading), while others involve flexible time on timed tests, or flexible settings. For example, in 1994, 14,994 seniors took a special edition of the SAT that has been offered over the years for students with disabilities. This version of the SAT allows large-block answer forms, having an aide read questions and record answers, or using extended time to complete the test. The most thoroughly investigated accommodation is that of handicapped students who are given extra time to complete the GRE or SAT. The conclusion of such research has been that this accommodation may be a source of test score comparability problems (that is, extended time administrations may lead to over-estimation of ability or achievement -- see Willingham et al. 1988). The accommodation issue must also be examined in the context of the move from paper and pencil to computerized testing. For a recent broad survey of testing accommodation issues, see Thurlow, Ysseldyke & Silverstein, 1993.

[21] Hambleton (1993) discusses the problems of translating achievement tests as well as some of the ways major problems might be addressed. See also Secada (1994).

learning environment, a school system, an economy, and a society dominated by the English language, use of a native language test may not be predictive of later outcomes.

About three quarters of LEP students are speakers of Spanish. There are important dialect differences within Spanish that must be taken into account in translation. The remaining quarter of the LEP student population comprises speakers of numerous other languages. In both HS&B and NELS:88, a Spanish-language questionnaire was made available to students in order to maximize the number of LEP students in the study. In both HS&B and NELS:88, the same questions were posed to students who took the English and Spanish version, and data from both groups of questionnaire takers were merged in the overall data set. However, very few students elected to complete the questionnaire in Spanish. In the 1980 HS&B base year, for the senior cohort, only 13 Spanish questionnaires were completed compared to 28,214 questionnaires in English; the 13 Spanish questionnaire completions seems a low number given 3,177 participating self-identified Hispanics, a substantial subset of whom may be presumed to have been LEPs . The 1980 sophomore cohort saw more Spanish questionnaire completions -- 43 students -- though this still represents only a little more than one tenth of one percent of the sample and one percent of Hispanics (there were 29,987 participants, 3,521 of whom identified themselves as Hispanics). Similarly low proportions of Spanish-language student questionnaire completion (around a tenth of a percent of the total sample and less than one percent of Hispanics) were recorded in NELS:88 in 1990 and 1992. In contrast, 22-23 percent of Hispanic parents chose the Spanish questionnaire option when the NELS:88 parent survey was conducted in 1988 and 1992. In the base year, of 2,496 participating Hispanic parents, 575 completed the Spanish version of the parent instrument. In the second follow-up, of 1,710 participating Hispanic parents, 373 chose to complete the Spanish-language parent questionnaires The method by which a choice of English or Spanish was presented to parents was somewhat different in the parent and the student surveys.[22]

While it is to be expected that some Hispanic LEPs have limited proficiency in the reading and writing of Spanish as well, and that many more parents than students would require a native-language instrument, one may also at least speculate whether there are other reasons for these low numbers of Spanish-language student questionnaire completions in HS&B and NELS:88. It is possible, for example, that Spanish-speaking students in an English-language setting are hesitant to publicly draw attention to their language status, or that interviewers who are not Spanish speakers have a difficult time communicating with such students and conveying a sense of their questionnaire options. Hence it is also possible that some students who could much better have answered in Spanish chose not to do so. As a topic for future methodological research, we recommend that means of conveying the option to complete a questionnaire in English or another language be studied to determine the best method to encourage those who would benefit from using a non-English version of the instrument to do so.

It should also be noted that an important function may be filled by having a Spanish questionnaire, regardless of how many students actually opt to complete it, and regardless of whether most students who would benefit from it actually use it. The fact of the availability of a Spanish language student questionnaire removes from school personnel a major reason (or excuse) for excluding Hispanic LEPs.

---

[22] In the base year, Hispanic ethnicity was gathered at the time of student sample selection. In the initial mailing of questionnaires to parents, both English and Spanish questionnaires were mailed to parents of Hispanic sample members so that an Hispanic parent could complete the questionnaire in the language with which the parent was more comfortable. In addition, follow-up contacts utilized bilingual interviewers; Spanish-speaking interviewers were trained to administer the questionnaire over the telephone in Spanish when necessary.

Chapter 2: Characteristics of Base Year Ineligibles

Proportions reported below in tabular and graphic presentations describing the characteristics of excluded 1988 eighth graders are weighted estimates, that is, they reflect expansion of counts from sample data to the full population of 1988 eighth graders as defined in the sampling plan, and take into account the higher rates at which Asian and Hispanic students were sampled into NELS:88 (see *NELS:88 Base Year Data File User's Manual*, p.18; and *NELS:88 Base Year Sample Design Report* for a complete description of the NELS:88 sample design).

2.1 Reason for Exclusion

Table 1 and Figure 1 below show breakdowns of the base year ineligible sample by reason for exclusion, in weighted percents. About a fifth of the excluded students were ineligible because they were limited in their English language proficiency, while nearly two thirds were ineligible owing to a mental disability. Only 6 percent were excluded by virtue of a physical disability. (Sex and race breakdowns are also of note and are discussed at length below, in sections 2.2 and 2.3 respectively.)

Table 1
Characteristics of the Base Year Expanded and Ineligible Samples: as Percentage of Expanded Sample and as Percentage of BYIs

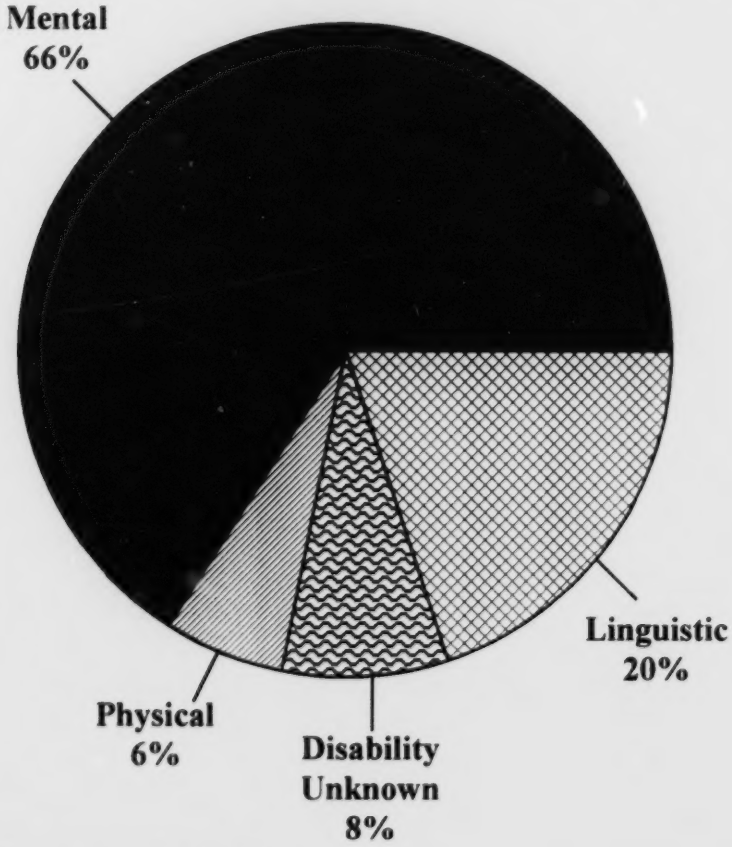
	% of Expanded Sample	% of BYIs
NOT EXCLUDED	95.3	0.0
EXCLUDED -- BARRIER:		
Physical	0.3	6.0*
Mental	3.1	65.8
Language	0.9	20.3
Unknown	0.4	8.0
RACE		
Asian	3.5	7.3
Black	13.8	15.7
Hispanic	10.9	18.9
White	68.6	47.4
American Indian	1.6	2.0*
Other	1.6	8.6
SEX		
Male	50.8	63.5
Female	49.2	36.5

Table 1
Characteristics of the Base Year (1988)
Expanded (Full) and Ineligibles Samples: Summary
(Continued)

	% of Expanded Sample	% of BYIs
RACE x SEX		
Asian-Female	1.7	3.0
Asian-Male	1.9	4.4
Black-Female	6.7	5.3
Black-Male	7.1	10.5
Hispanic-Female	5.6	9.1
Hispanic-Male	5.2	9.8
White-Female	33.8	15.6
White-Male	35.8	31.8
American Indian-Female	0.8	0.4
American Indian-Male	0.8	1.6
Unknown-Female	0.7	3.2
Unknown-Male	0.9	5.4

SOURCE: **NELS:88 Second Follow-Up Survey (1992), National Center for Education Statistics, U.S. Department of Education.**

Figure 1:
Base Year Ineligibles by Reason for Exclusion
Weighted Percents



SOURCE: National Education Longitudinal Study of 1988: Second Follow-up, NCES, US ED.

The two disability categories (mental and physical) used in NELS:88 are overly broad. Table 2 displays the more fine-grained disability categories of the Office of Special Education Programs: specific learning disability; speech or language impairment; mental retardation: serious emotional disturbance; hearing impairment; orthopedic impairment; other health impairments: visual impairments; multiple disabilities; deaf-blindness. It is probable that the distribution of specific handicaps within the total disabilities sample of NELS:88 (students with disabilities who were included, plus those who were excluded) is highly similar to the proportions in Table 2. The breakdowns in Table 2 provide information that is crucial for three determinations. First, it provides a basis against which to compare the NELS:88 distribution of subpopulations. (This comparison is qualified, however, by the fact that excluded 1988 eighth graders encompassed only a select subset of eighth graders with disabilities receiving special education services.) Second, the table shows the rarity of disability subpopulations. Some, such as students with visual impairments, are sufficiently rare that extraordinary levels of oversampling would be required to obtain analyzable numbers. Sample sizes for national studies rarely exceed a thousand schools, an important limit if within-school oversampling is to be pursued. (Some rare populations of course may be clustered in certain specialized or regular schools, providing additional opportunities for building a sample.) Third, by depicting the various subcategories of students with disabilities, Table 2 suggests the range of testing accommodations that might have to be made to maximize inclusion. (For example, accommodations for students with an emotional disturbance may entail one on one administration or split sessions, while adaptations required for a hearing impaired student might include using sign language to provide instructions.)

Table 2
Public School Children 0 to 21 Years Old Served in Federally Supported Special Education Programs in 1987-88 by Type of Disability

	Number served (thousands)	Percentage Distribution	N Served as % of Total Enrollment
All disabilities	4,447	100.0	11.1
Specific learning disabilities	1,928	43.4	4.8
Speech or language impairments	953	21.4	2.4
Mental retardation	582	13.1	1.5
Serious emotional disturbance	373	8.4	0.9
Hearing impairments	56	1.3	0.1
Orthopedic impairments	47	1.1	0.1
Other health impairments	45	1.0	0.1
Visual impairments	22	0.5	<0.1
Multiple disabilities	77	1.7	0.2
Deaf-blindness	1	<.05	---
Preschool disabled	363	8.2	0.9

SOURCE: **Modified Presentation of Table 50, *The Digest of Education Statistics 1992*. Statistics source is U.S. Dept. of Education, Office of Special Education and Rehabilitative Services, Annual Report to Congress.**

2.2 Sex

Are there significant gender differences in the likelihood of being excluded from national education data collections? To the extent that some critical classification groups are disproportionately excluded, bias in subgroup estimates is a possible consequence.

For eighth grade LEP students, the presumption would be that male to female ratios in the excluded student group would be fairly close. (Ratios might differ for a later grade cohort if there was a large differential by sex in the dropout rate).

For students with disabilities, however, external sources suggest that male and female representation could differ. Data collected by the U.S. Department of Education's Office of Civil Rights shows that in the 1988 school year, 64.7 percent of students in special education programs for learning disability, speech impairment, mental retardation, or serious emotional disturbance, were male, and 31.6 percent female. Some 47.5 percent of the total had learning disabilities. The sex breakdown for learning disability as a proportion of the total was 33.9 percent for males and 14.2 percent for females.[23] Keeping in mind that the NELS:88 excluded students comprise only a subset of NELS:88 students receiving special education services (since other students with IEPs were included in the sample), it is nevertheless of interest to examine the ratio of males to females excluded owing to disabilities.

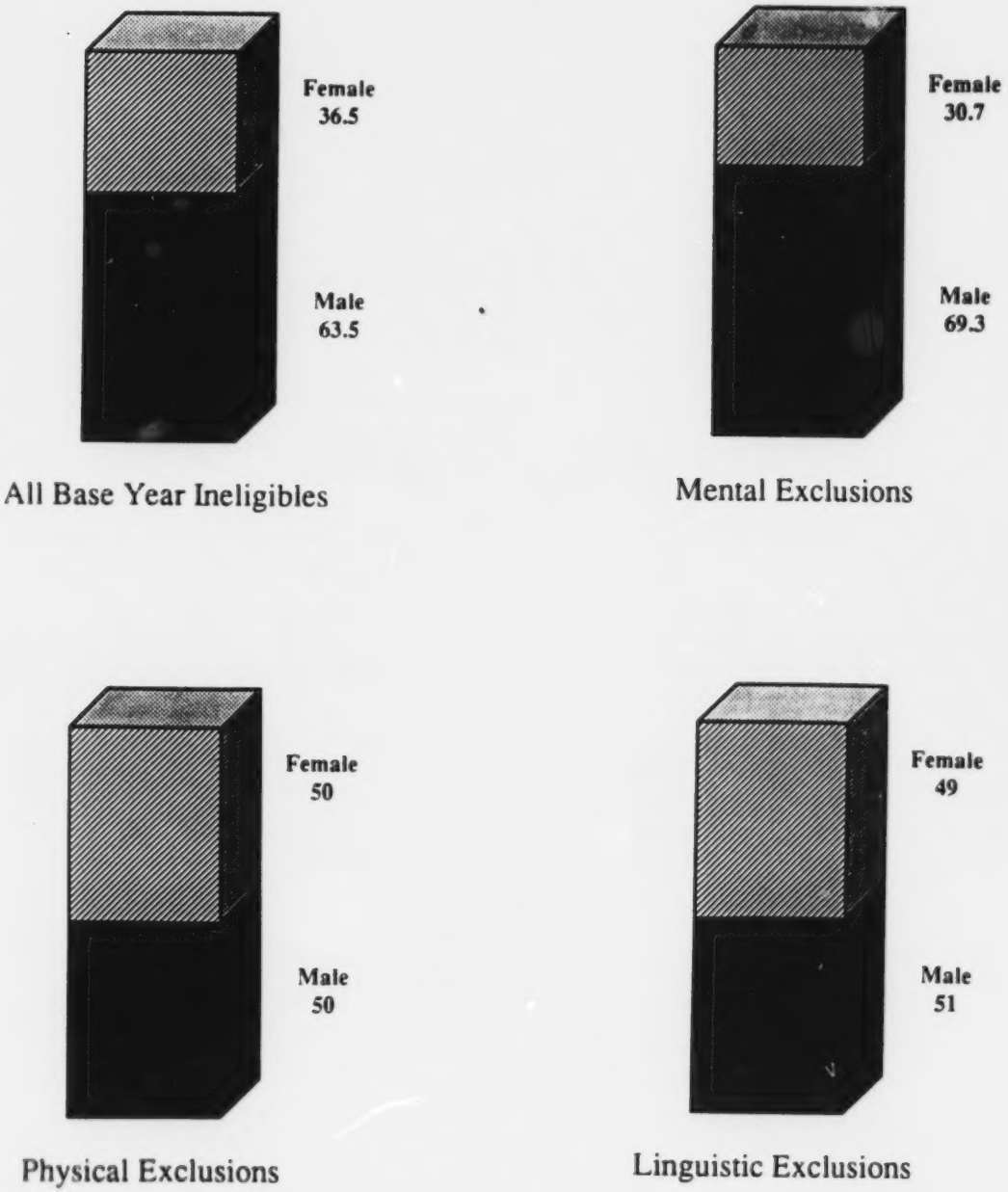
Table 1 indicates that some 63.5 percent of the excluded eighth grade cohort members were males, and 36.5 percent females. Table 1 also reports total sample and ineligible sample sex proportions in cross-classifications with race.

Reason for exclusion by sex. The distribution of males and females, however, varies by type of exclusion, as is illustrated in Figure 2. Figure 2 displays weighted estimates of the male/female proportions for the excluded students as a whole (36.5 percent female, versus 63.5 percent male); for eighth grade cohort members excluded in 1988 for reasons of a mental disability (30.7 percent female and 69.3 percent male); for those excluded for reasons of a physical disability (50 percent female, 50 percent male); and for those excluded because of limited English language proficiency (49 percent female, versus 51 percent male).

[23] This information is taken from supplemental Table 45-4, p.304, *The Condition of Education, 1994*.

Figure 2:
Exclusion Status by Gender Breakdown

Weighted Percents



SOURCE: National Longitudinal Study of 1988: Second Followup, NCES, US ED.

2.3 Race/ethnicity

Again, subgroup estimates may be biased if some subgroups are disproportionately excluded from assessments and surveys. Because of differential immigration rates by different groups, one would expect to see Hispanics and Asians dominate the excluded LEP group in the NELS:88 sample.

In terms of students with disabilities, Office of Civil Rights data show some differences in race/ethnicity proportions when the general student population and students in special education programs are compared. Overall, in 1990, proportions of students receiving special education services within each major racial/ethnic group were as follows: 11.3 percent of blacks, 10.8 percent of American Indians, 9.5 percent of whites, 8.2 percent of Hispanics, and 3.7 percent of Asians.

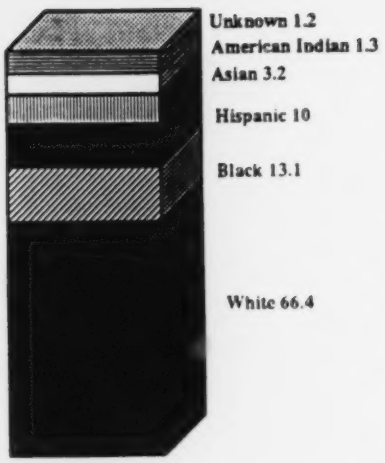
Within special education subcategories, there is in particular some overrepresentation of blacks in the mental disability conditions.[24] Again, it should be kept in mind that NELS:88 excluded students do not encompass all students with disabilities in the NELS:88 sample, or all LEP students, since substantial numbers of such individuals were included in the base year.

Table 1 (above) shows the weighted proportions by race and Hispanic ethnicity for the base year excluded students, as well as the proportions falling into each racial and ethnicity group for the entire (eligible and ineligible) eighth grade cohort sample. Some 7.3 percent of the excluded eighth grade cohort was Asian, some 15.7 percent black, 18.9 percent Hispanic, and 47.4 percent white. Because not all members of the ineligible sample were located, the category for "other" (which includes unknown race) is comparatively high, at 8.6 percent.

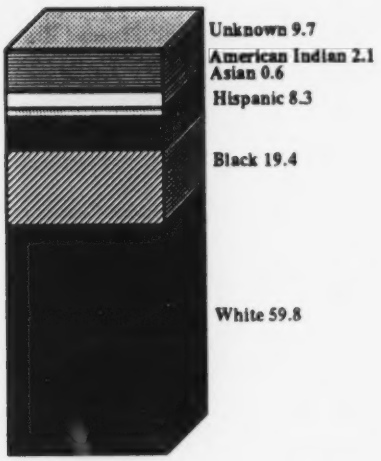
Mental disability by race/ethnicity breakdowns. Figure 3 (below) looks at the racial/ethnic breakdown within the largest exclusion category, mental disability. It shows that of those excluded for this reason, about 59.8 percent were white, 19.4 percent black, 8.3 percent Hispanic, and 0.6 percent Asian. (Those of unknown, or other race, or rare populations such as American Indians and Alaska Natives, have been excluded from this calculation).

[24] See supplemental table 45-4, *The Condition of Education, 1994*, for 1986, 1988, and 1990 statistics. Also OCR 1993.

Figure 3:
Racial/Ethnic Breakdown for Eligibles and BYIs with a Mental Disability
1992 Weighted Percents



Base Year Eligibles



Base Year Ineligibles with Mental Disability

SOURCE: National Longitudinal Study of 1988: Second Followup, NCES, US ED.

2.4 1992 Enrollment Status: students and dropouts; grade progression

Students and Dropouts. Table 3 addresses the 1992 school enrollment status of the excluded 1988 eighth graders. Only 62 percent of the eighth grade cohort 1988-ineligible population was still in school four years later. Some 7.6 percent, however, were taking some sort of educational program as preparation for a GED or other alternative educational credential. These statistics -- the proportion of eighth graders who were dropouts four years later -- do not constitute a cumulative dropout rate (that is, an event tally of all students who ever dropped out), which would be somewhat higher, insofar as some students who dropped out of school between eighth and tenth grade presumably would have returned by 1992, but would have fallen behind the grade progression of those eighth grade cohort peers who remained in the modal sequence.

The NELS:88 studies of base year excluded students were designed to gather information about enrollment status that would permit overall eighth grade cohort dropout rates to be adjusted to reflect the status of the entire cohort, ineligible and eligible alike. The ineligibles followback studies were not designed to provide national dropout rates for subgroups of excluded students, or for the sum total (excluded and included) of LEP and IEP students. However, the U.S. Department of Education's National Longitudinal Transition Study of Special Education Students (NLTS), which began in 1987, was designed to provide national school leaving estimates for students with disabilities, and confirms key features of the NELS:88 descriptive statistics reported in Table 3. In particular, students with disabilities drop out at a high rate. NLTS data show that overall, students with disabilities register a cumulative ninth through twelfth grade dropout rate of 29.9 percent (a further 8 percent of students with disabilities leave school prior to ninth grade). The NLTS also shows the highest grade nine to twelve dropout rates for students classifiable under the NELS:88 scheme as having a mental barrier to participation, and comparatively lower rates for those with a physical disability. For example, the cumulative (grade nine to twelve) dropout rate for the learning disabled is 28.5 percent, for the mentally retarded 29.9 percent, and for the emotionally disturbed 48.1 percent. On the other hand, the cumulative dropout rate for the visually impaired is 12.1 percent, for hard of hearing 14.9 percent, for the deaf 11.3 percent, and for the orthopedically impaired 13.5 percent. (For further details see Wagner, Blackorby, Cameto, & Newman, 1993, Table 5-1, p. 5-2 or OSEP 1994, Table 3.9, p.99; on dropout issues pertaining to students with disabilities, see also Wolman, Bruininks & Thurlow, 1989.). In terms of grade progression, those special education students who remain in school are less likely than their non-special education peers to stay within the normal grade progression and complete school on time (Hayward and Thorne, 1990).

In terms of Hispanic LEPs, who dominate national LEP statistics by constituting three quarters of this population, there is also evidence of high national dropout rates and high rates of repeating a grade. In NELS:88, among LEP students who were classified as eligible in the base year (and, presumably, were in general the more English-proficient of the LEPs), 65 percent received a high school diploma in 1992 compared to an overall rate for the eighth grade cohort of 82 percent.[25] The status dropout rate for Spanish-speaking Hispanics (32.1 percent) is three times the overall rate (11 percent) and the rate increases

[25] See Green, P.J., & Scott, L.A., 1995, *"At-Risk" Eighth Graders Four Years Later*, Statistics in Brief Series, Washington, D.C., NCES.

as reported English-language facility decreases.[26] Current Population Survey data for 1989 distinguish Spanish from other language LEPs. These data indicate that 34.7 percent of all children were enrolled below modal grade, that 37.7 percent of all LEPs were below modal grade and that this was true of 42.1 percent of Hispanic LEPs and of 32.8 percent of Asian LEPs.[27]

Grade Progression. Table 3 also indicates that of those who were still in school, roughly 58 percent of excluded 1988 eighth graders were 1992 seniors while the other 42 percent were not 1992 seniors, that is, had fallen behind the 1988 eighth grade cohort's modal progression through school. Figure 4 examines 1992 in-sequence and out-of-sequence 1988 ineligibles by their reason for exclusion from the sample. While 80 percent of those excluded because of a physical disability were 1992 seniors, 55 percent of those excluded for a mental disability were 1992 seniors, as were 67 percent of those excluded owing to limitations in their proficiency in the English language.

Table 3
1992 School Enrollment Status of the Base Year (1988) Ineligible Sample: Summary

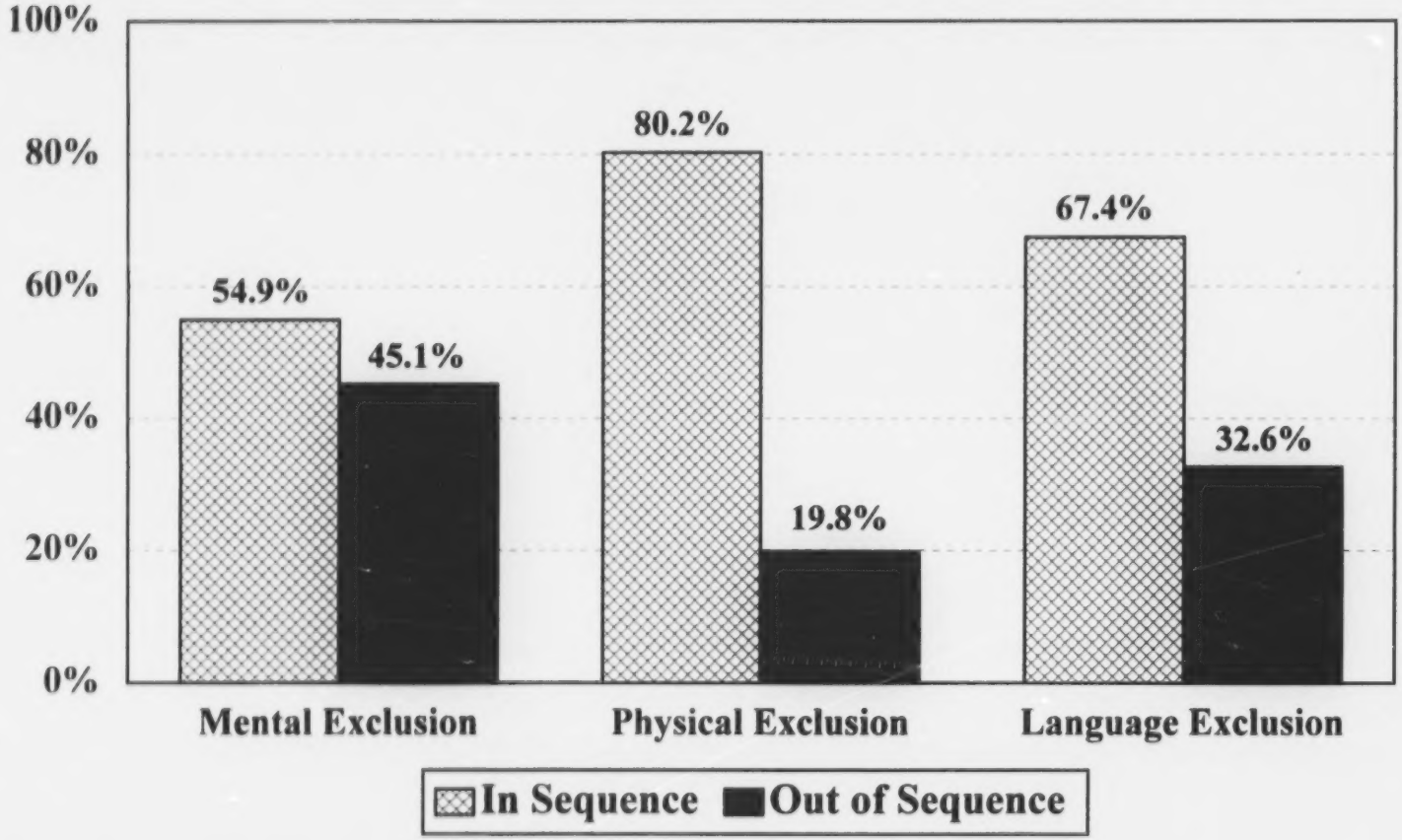
	Expanded Sample	BYIs
ENROLLMENT		
In School	83.3	62.4
Dropout	11.6	30.0
Alt. Completer	5.2	7.6
IN SEQUENCE		
Yes	80.0	57.6
No	20.0	42.4

SOURCE: NELS:88 Second Follow-Up Survey (1992), National Center for Education Statistics, U.S. Department of Education.

[26] See McMillen, M., Kaufman, P., Hausken, E. & Bradby, D. *Dropout Rates in the United States: 1992.* Washington, D.C., National Center for Education Statistics. See especially Table 12, which shows that the status dropout rate for Hispanics who report that they speak English very well is 17 percent and for those speaking English well it is 30.4 percent; for those who report that they do not speak English very well, the rate is 61.8 percent and for those who report no ability to speak English it is 83 percent.

[27] See *The Condition of Education, 1992*, p.22.

Figure 4:
Base Year (1988) Ineligibles in 1992:
Grade Progression by Reason for Exclusion



SOURCE: National Education Longitudinal Study of 1988:
Second Follow-up, NCES, US ED.

Chapter 3: Eligibility Status of Excluded 1988 Eighth Graders in 1992

A major question is what proportion of the base year ineligibles later became eligible. It is not possible to know definitively what proportion of students whose classification changed truly represent status changes as contrasted to reclassifications that reflect changes in the individuals making eligibility determinations and a sharpening of the eligibility criteria themselves. Nevertheless, the magnitude of change supports the notion that it would have been a major error to not have returned to the excluded student sample in the subsequent rounds, for the majority of excluded students were eventually brought into the NELS:88 sample.

Table 4 summarizes, as a weighted proportion of the total eighth grade cohort, the 1992 eligibility status of 1988 excluded eighth graders overall and by sex and by race/ethnicity. It shows that 4.7 percent of 1988 eighth graders were excluded, with 1.7 percent of the cohort still excluded in 1992 and the remainder of the base year ineligibles registering as another 0.5 percent of the cohort who proved unlocatable.

Table 5 summarizes outcomes for 1988 eligibles in terms of sample Ns and unweighted proportions. It shows that for the 1988 excluded sample in 1992:

- 57 percent were now eligible, 32 percent remained ineligible, and for 11 percent (unlocatables), status had not been ascertained
- 71 percent of the LEP excluded students had become eligible
- 50 percent of IEP students with a mental impairment had become eligible
- 57 percent of IEP students with a physical disability had become eligible

It should be noted that most reclassifications took place in the first follow-up.[28] Caution should be exercised in interpreting data for students with physical disabilities, given their sparse representation (N = 23) in the sample. The large number of out of scope 1988 language barrier students (25 out of 202) is to be attributed to out-of-the-country status in 1992.

Table 6 reports sample sizes for excluded students and their weighted population proportion within the total eighth grade cohort, reflecting 1992 eligibility status of 1988 excluded eighth graders by their reason for exclusion. Table 6 shows that of the three percent of the eighth grade cohort excluded because of a mental disability in 1988, by 1992 1.4 percent were eligible, another 1.4 percent remained ineligible and .2 percent were unlocatable (hence their status could not be ascertained). For the one percent of 1988 eighth graders excluded because of a lack of proficiency in English, by 1992 .7 percent were eligible, .1 percent remained ineligible, and .2 percent unlocatable. These proportions, which are *population estimates* derived from use of case weights, differ somewhat from those in Table 5, which reports on the sample.

Figure 5 displays the 1992 eligibility status of 1988-ineligible eighth graders, broken out by reason for exclusion. In other words, Figure 5 illustrates graphically the data reported in Table 5.

[28] See Ingels, Scott, Rock, Pollack & Rasinski, 1994, Table 7.5, for details.

Table 4
Expanded Sample 1992 Eligibility Status Summary, in Percents

	1988 I 1992 E	1988 I 1992 I	1992 Unlocatable	1988 E 1992 E
TOTAL	2.5	1.7	0.5	95.3
SEX				
Male	1.5	1.2	0.3	47.8
Female	1.0	0.5	0.3	47.5
RACE				
Asian	0.2	0.1	0.0	3.2
Hispanic	0.4	0.3	0.1	13.0
Black	0.6	0.1	0.2	10.0
White	1.1	1.0	0.1	66.4
American Indian	0.1	0.2	--	1.5

Key: E = eligible
I = ineligible

Note: Eligibility status could not be ascertained for unlocatable sample members.

SOURCE: **NELS:88 Second Follow-Up Survey (1992), National Center for Education Statistics, U.S. Department of Education.**

Table 5
Summary of Final 1992 Statuses for 1988
Excluded Students ("Followed-back" Sample)
in Unweighted Percents

Reason for 1988 Exclusion	Eligible		Ineligible		Not Ascertained		Sample N
Language Barrier	125	71.0%	22	12.5%	29	16.5%	176
Physical Barrier	13	56.5%	9	39.1%	1	4.3%	23
Mental Impairment	166	50.3%	140	42.4%	24	7.3%	330
Unknown Reason	30	54.5%	15	27.3%	10	18.2%	55
TOTAL	**57.2%**		**31.8%**		**11.0%**		**584**

(Excludes cases sampled in error and those out of scope [dead or out of country] for 1992 round).
(Owing to rounding, rows may not sum to 100 percent).

Table 6
Expanded Sample 1992 Eligibility Status
by Type of Exclusion

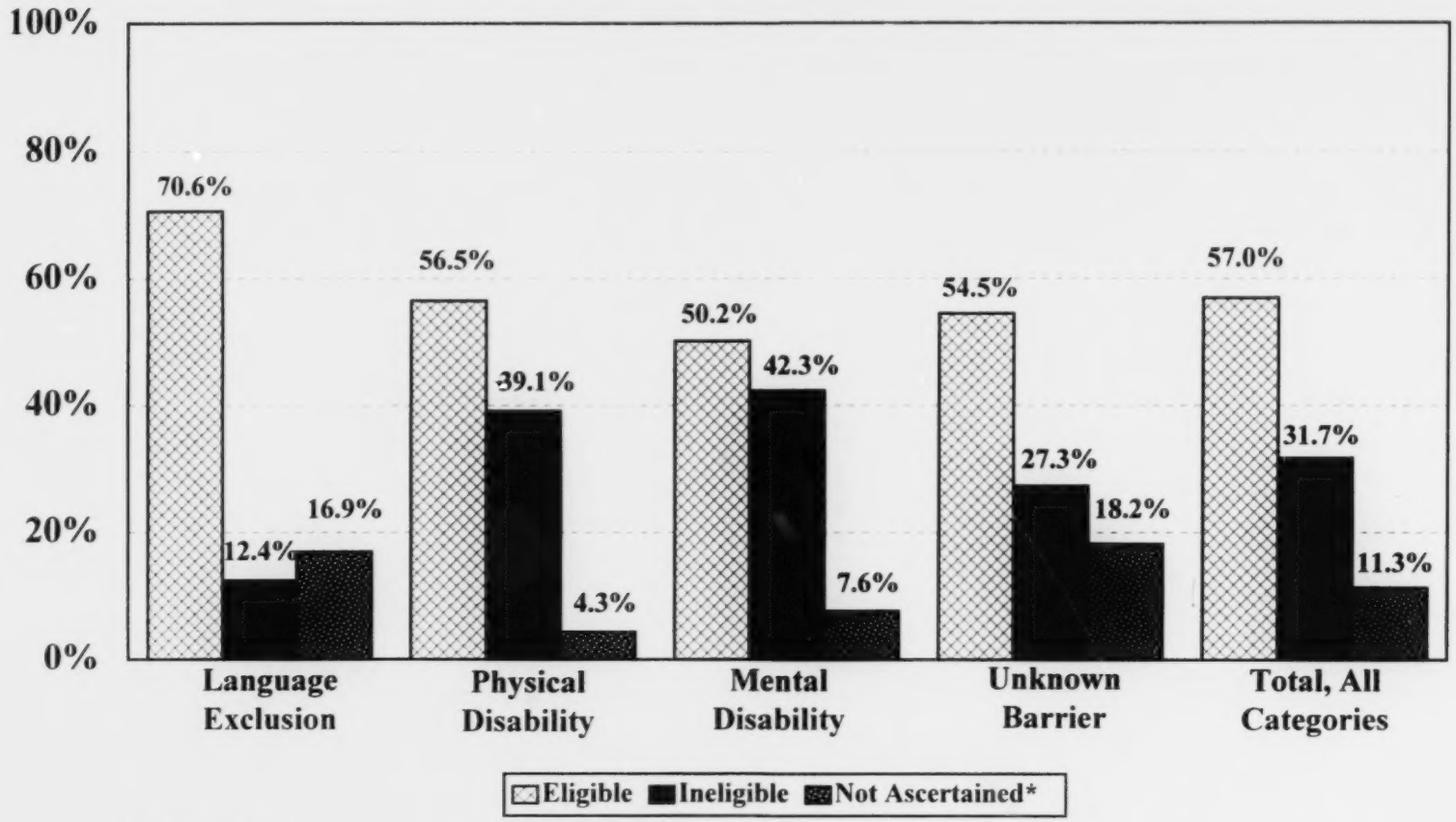
Barrier	1988I 1992E		1988I 1992I		1992 Unlocatable	
	N	%	N	%	N	%
Physical	13	0.1	9	0.1	1	0.0
Mental	166	1.4	140	1.4	24	0.2
Language	125	0.7	22	0.1	29	0.2
Unknown	30	0.2	15	0.1	10	0.1

Key: E = eligible
I = ineligible

Note: Eligibility status could not be ascertained for unlocatable sample members.

SOURCE: NELS:88 Second Follow-Up Survey (1992), National Center for Education Statistics, U.S. Department of Education.

Figure 5:
1992 Eligibility Status of 1988 Excluded Eighth Graders, by Reason of Exclusion



*Not Ascertained

SOURCE: NELS:88 Second Follow-Up Survey, NCES, US. ED.

Note: proportions reflect (unweighted) sample Ns.

Chapter 4: Other Potential Sources of Sample Undercoverage

There are seven ways in which a student may have failed to have a chance of selection into the NELS:88 base year. Three of these ways ("e", "f" and "g" below) are classic undercoverage problems. Another is a nonresponse problem ("a" below). The final three ("b", "c", and "d") involve population definition; one may by design for any number of reasons choose to exclude elements from the target population. However, we have argued that element "c", given the critical descriptive purposes of NELS:88, should not have been excluded from the target population (but only from the cognitive assessment, or in some instances student questionnaire, component of the study). If this is so, and the base year eligibility rules misspecified the target population, then "c" may be viewed as an instance of coverage error as well.

(a) First, if the student's school refused, that student had no chance of selection;

(b) Second, if the student's school was declared ineligible to participate, that student had no chance of selection;

(c) Third, though the selected school participated, the student was declared ineligible to participate, owing to disability or a lack of command of English;

(d) Fourth, the student was studying at home in 1987-88, or abroad, or in an ungraded program or school;

(e) Fifth, the student was unavailable (for example, was hospitalized, incarcerated, or was a migrant in transit during the period when sampling and sample updating took place);

(f) Sixth, owing to clerical error, the student did not appear on the correct roster or was misclassified.

(g) Seventh, the student's base year school had no chance of selection, because the sampling frame was inaccurate (for example, a student might attend a newly-opened school that had not yet been added to the school list from which the sample was drawn).

In addition to these seven ways in which a student may have failed to have a chance of selection into the base year, in the follow-up rounds when representative tenth and twelfth grade cohorts were generated,

(h) a student might have been omitted from the freshening process, for any of several reasons: roster or clerical error, failure to freshen on certain groups (for example, freshening was not performed on the base year ineligible sample; freshening was performed in schools, but not in institutional settings such as juvenile detention homes), and so on.[29]

[29] Through the process of sample freshening, 1990 sophomores and 1992 seniors who had no chance of selection into NELS:88 in 1988 (because they were not in the United States, or not in the eighth grade), are added to the data set. Most (but not all) of the same sources of potential undercoverage arise in freshening as in base year sample selection. As an example of the exceptions to this generalization, one is no longer dependent on a published universe list of schools and therefore vulnerable to omissions in the sampling frame, in that the schools at which freshening took place in 1990 and 1992 were the schools to which 1988 eighth graders dispersed, nor was any type of school ineligible in the follow-ups.

The focal point of this report is category "c", excluded students, whom we have discussed at length. We will now comment briefly on other possible sources of undercoverage.

School refusals (a). Since substitute selections replaced original selections that refused, this case does not pose a risk of undercoverage, but rather of potential bias.[30]

Ineligible schools (b). Virtually all schools in the fifty states and the District of Columbia that enrolled eighth graders in the 1987-88 school year were eligible for the study. However, Bureau of Indian Affairs (BIA) schools were categorically excluded from the 1987-88 school frame. Given that just over 1 percent of eighth graders at the time were American Indians and that 90 percent of American Indian students attend non-BIA schools, this exclusion should have a negligible impact on estimates, though it should be taken into account when considering results for the American Indian subgroup. Also excluded were special education schools for serving students with disabilities[31], area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas.

Excluded Students (c). Excluded students are a subclass of the ineligible students, specifically, those who were declared ineligible for reasons of mental, physical, or linguistic barriers to participation. While students who died, were part-time students primarily registered at another school, or who transferred out of the school prior to its Survey Day, were also declared ineligible, these categories of students should affect neither the representativeness of the sample nor estimates derived from it. The governing principle here is that each 1987-88 eighth grader should have one chance of selection into the NELS:88 sample, and only one. Part-time students with a primary registration elsewhere had a chance of selection into the sample at the site of their primary registration. Transfers out of the school were classified as ineligible, but sample representativeness was maintained by giving transfers into the school during the same period a chance of selection into the Base Year sample. However, the five percent of base year students with severe physical, mental or linguistic obstacles to participation had, as a class, no chance of selection into the sample; they were systematically excluded. Assuming their characteristics and behaviors to be in any essential way different from the norm, their exclusion will be a source of undercoverage bias in national estimates.

Home Study, Abroad, Ungraded (d). These students were deemed not to be part of the target population; the "ideal" target population was all eighth graders enrolled in schools in the fifty states and the District of Columbia. While students not enrolled in an American school but receiving an education at home or abroad were not eligible for selection into the base year, such students had a chance of selection into the study in 1990 or 1992, if their status had changed, that is, if they were in the tenth grade in a school in the United States in the 1990-91 school year or the twelfth grade in an American school in the 1991-1992 school year. Implicitly, students in ungraded programs (which, historically, have been

[30] Potential school nonresponse bias is analyzed in Spencer, Frankel, Ingels, Rasinski & Tourangeau, 1990.

[31] According to Office of Special Education figures reported in the *Digest of Education Statistics, 1992*, Table 51, 5.5 percent of special education students receive services in separate schools or residential facilities, while .8 percent are in a homebound or hospital environment. Not all of these individuals are in graded programs. Separate facilities tend in particular to be available for comparatively rare populations such as individuals with severe visual or hearing impairments, and for emotionally disturbed students whose presence might impede regular classroom activities. Most students who are doubly physically disabled by being both deaf and blind are educated in special facilities.

the service vehicle for the mentally retarded[32]) are excluded, since NELS:88 is a grade cohort, not an age cohort, and such students will not appear on an eighth grade school roster.

Temporarily unavailable (e). Rostered students undergoing prolonged hospitalization or institutionalized or otherwise unavailable were extremely rare in the base year. However, in the NELS:88 follow-ups, substantial numbers of limited English proficient (LEP) students, particularly in the northeast and on the west coast, had left the country at the time of data collection. Such students are regarded as temporarily out of scope in NELS:88, and subject to re-survey should they have returned to the United States at the time of the next data collection. Migrant students may be a group that is particularly hard to represent within a school-based sample. Generally the most stable period for sampling this group, that is, the time at which they are likely to be at their home base school and not in transit, is early in the calendar year. Because of the small size of the migrant student population (about seven tenths of one percent of public school enrollment, per Henderson, Daft and Gutmann, 1989), some under-representation of this group would not pose a large risk of biasing national or subnational estimates. The sample update procedure, by which new transfers into the school before survey day were given a chance of selection into the base year, should be expected to minimize any underrepresentation of this group. Students receiving migrant education services are not separately identified on the NELS:88 database.

Misclassifications and omissions (f). A small number of cases have been removed from the NELS:88 sample owing to later discovery that the student was in a grade other than grade 8 at the time of sampling, and appeared on an eighth grade roster in error. Presumably some number of cases that should have been listed on eighth grade rosters did not appear. While the number of such cases is likely to be quite small, there is no way to be certain precisely how many eighth graders may have been omitted from school listings.

However, omission of **transfer** students is a quantifiable problem. NELS:88 followed essentially the same procedure for dealing with transfer students as did High School and Beyond (HS&B) in 1980. School rosters were submitted and an initial sample drawn in the autumn. To adjust the student sampling frame for student attrition and change in the eighth grade population of the sampled school, NORC conducted a sample update seven to ten days prior to the school's scheduled survey session. The NORC survey representative went over the sample list with the school coordinator to ensure that all sampled students were still eligible, and that transfers-in--that is, any student who had joined the eighth grade class of the school between the time of the original sampling and the time of the update--were added to a supplementary roster from which additional students would be selected. Selections for inclusion in the sample were based on the same set of computer-generated random numbers used to select the original sample.

Given that mortality and dropout rates are very low for eighth graders, in theory, there should be a rough parity in the number of selected students lost to transfer and the number selected into the sample from the pool of transfers in. Overall, around four percent of the NELS:88 original sample had transferred out by survey day, but the replacement rate was around two percent, half the expected percentage (Ingels, Rizzo and Rasinski, 1989). This experience is not peculiar to NELS:88. For example, for the National Assessment of Educational Progress (NAEP) Trial State Assessment in 1990, Spencer (1991, p.6) reports that 4.9 percent of the students withdrew from the sample but the supplemental sampling procedure added only 2.9 percent to the sample. Thus there was a 40 percent undercoverage

[32] At present "91 percent of elementary and secondary public special education students are in graded classes (or placements)" according to the National Council on Disability (1993).

of transfer students. The reason for undercoverage of transfers in the base year of a longitudinal study or in a cross-sectional study would appear to be that while all transfers out will be identified successfully (any missed outward transfers at the sample update will be identified when the no-shows at survey day are investigated), school records are not always sufficiently accurate and up to date to provide definitive lists of all students who have transferred in since a certain date.

Inaccuracy of sampling frame (g). Schools in 1987-88 that had just come into existence in the previous several months were not represented on the lists from which the sample was drawn. There may be very modest undercoverage of very small private schools, particularly those of recent origin.[33]

[33] For an assessment of the adequacy of 1986 list frame data for school sampling for the 1987-88 school year, see Kaufman, 1991, which reports on the experience of the NCES Schools and Staffing Survey (SASS). SASS employed a dual frame approach to private schools, with an area frame used to find schools missing from the list frame, thereby compensating for undercoverage in the list frame.

Chapter 5: Conclusions and Recommendations

5.1 The Need for Greater Inclusiveness

The movement toward greater inclusiveness in assessment and research programs is driven not just by the desire for greater statistical precision (such as, minimization of undercoverage bias) but also by a wider policy context that endorses inclusiveness of instruction and assessment as a desirable goal. Indeed, a major policy theme in the discussion both of IEP and LEP students is that of their integration into regular classrooms. Given increased emphasis in recent years on educational standards and outcomes (particularly test scores) and stress on the goal of integrating assessment into instruction, integration of special needs students into regular classrooms has been seen as arguing also for their integration into assessment and research programs, particularly those that are seen as driving current school reforms.

Education longitudinal studies should be designed so that they will map the school careers of students with mental or physical disabilities or limited English proficiency in a manner that facilitates a clear evaluation of such children's integration and progress. Normative conceptions of such progress and critiques of past and current special education, bilingual education, and English as a second language practice are needed to ensure that critical indicators are incorporated into new surveys.[34] While NCES longitudinal studies have stressed cognitive development, a more holistic approach, in which domains such as socioemotional development are assessed as well, is desirable and will be pursued in new studies such as the Early Childhood Longitudinal Study (ECLS). To study socioemotional development, state-of-the-art conceptions and measures of social adaptation as well as indicators of conduct disorder and other adjustment disorders must be incorporated and analytic methods devised to model trajectories of socioemotional development and the impact of school and classroom practices on these trajectories. The special strengths of NELS:88 and like data sets for conducting disabilities and other special populations research (as well as correctable weaknesses in the NELS:88 approach) are outlined in Hodapp & Krasner (1995).

Moreover, longitudinal studies with an assessment component may be able to play a particularly important role in helping to assess the predictive validity of innovative strategies for assessing IEP and LEP students. Studies such as NELS:88 are rich in home background variables, as well as school context and process measures, and most importantly, permit assessment results to be viewed in the light of subsequent educational, social and occupational outcomes.

Students with Disabilities. The 1975 Individuals with Disabilities Education Act specifically requires that students with disabilities be educated in the "least restrictive environment possible." Since 1975, there has been a notable movement toward "full inclusion" -- that is, teaching children with disabilities together in the same classrooms with other children.[35] Although not without its critics, and though the prospects of successful inclusion may depend in part on the severity and type of disability, there is a decided trend toward greater integration. By the 1991-92 school year when most members of

[34] In the area of special education, particular note should be taken of the work of the National Center on Educational Outcomes in the area of developing a comprehensive system of educational outcome indicators for students with disabilities. (See OSEP 1994, pp.131-152).

[35] For a discussion of assumptions that underlie arguments about inclusion, counter-assumptions, and a summary of current debates and practices, see *The Harvard Education Letter*, July-August 1994, X(4).

BEST COPY AVAILABLE

the NELS:88 eighth grade cohort were seniors, some 34.9 percent of the nation's students with disabilities were served in regular classrooms during at least 80 percent of the school day (OSEP, 1994, p. 14), the highest proportion ever. A resource room was the primary educational environment of 36.3 percent, 23.5 percent were in separate classes, 3.9 percent in a separate school, and 1.4 percent in a residential facility, homebound, or hospitalized. This trend away from segregated special education underlines both the desirability and possibility of including more such students in national assessments and research programs.

Students with Limited English Proficiency. The number of language minority, and LEP, students continues to grow. Immigration is at an all time high, with nearly one million new entrants annually. Data from the 1990 decennial census indicate that between 1980 and 1990 the number of Americans speaking a foreign language at home increased from 23.1 million to 31.8 million, of whom 17 million are Spanish-speaking (U.S. Census Bureau, 1993). While there has been no analogue to the special education inclusion movement, based on philosophical principles and backed by legislative imperatives, toward including LEP students in the same classrooms as non-LEP students, some school systems have implemented intensive English as a second language programs in which LEP students receive academic instruction in mainstream classrooms as well as language instruction in separate English as a Second Language (ESL) classrooms.

Transitional bilingual programs are more common, in which students begin learning all subjects in their native language while also participating in ESL programs for one or more periods during the day. Once students are considered proficient in English, they are moved into monolingual classes with English-speaking students. While there are debates about the instructional role of the native language with differences particularly concerning the rate and manner in which students should be integrated into regular classrooms, and different program orientations (for example, immersion versus transition) reflect these debates, there is widespread agreement that the main goal is to eventually make LEP students proficient in English and place them in English language classrooms (Hakuta & Garcia, 1989). From the point of view of a longitudinal research agenda, this strongly suggests the need to assess and, over time, reassess LEP students' *competency in English and progress in learning the English* language.

From the point of view of assessing various curricular areas such as science, mathematics, and social studies and humanities knowledge, the problem for achievement growth analysis in a longitudinal study is that LEP students will normally by degrees cross over into competence in these areas in English. Achievement scores are highly language-dependent, a factor compounded by the fact that most content instruction received by LEP students is in English (see, e.g., Moss & Puma, 1995, p.i-7, p.i-4). Developing a single standardized achievement growth measure that would validly assess the content achievement of LEP students is an enormous challenge that requires much further research. To the extent that native language translations of assessments are to be pursued with the hope of putting them on the same scale with English language test batteries, an IRT model should be used to determine, independently of ability differences in the examinee samples, the equivalence of items (Hambleton, 1993).

A more modest goal for longitudinal studies in the near-term may be to institute a mix of strategies along the following lines: (1) in the baseline, assess the student's degree of competence in English, and reassess English language competence in each follow-up; (2) when competence is achieved (based on tested English literacy level), assess the student in English; (3) in the baseline only, obtain some control measures based on native language assessment, e.g. a measure of vocabulary (there are Spanish and French versions of the PPVT, for example) and numeracy (for say four and five years olds, assessments such as the Test of Early Mathematics Ability could relatively easily and unproblematically be translated into Spanish or other languages). For older children literate in their native language, a

writing assessment in the native language would constitute valuable baseline information. The possibility of administering nonverbal measures also should be explored.

Another promising avenue, but one which requires a great deal of further research and development effort, is that of modifications in English-language assessments for LEP students to minimize the impact of the interconnection between language and content by decreasing the English language "load". Test instructions could be simplified and improved, and items could be modified through syntactic or semantic simplification (while maintaining the level of item content difficulty or conceptual complexity), or special assistance provided (e.g., dictionaries or glossaries). In addition, language contextualization instructional techniques such as scaffolding and sheltered instruction, could be applied to assessment (see Hafner 1994 and 1995 for an extended discussion, and Krashen 1991).

Other strategies include use of background variables to impute scores to LEP students, extended time, use of teacher reports, portfolios and various performance assessments, and so on; generally these are subject to limitations instanced in this report's discussion (below) of students with disabilities (although Hafner, 1995, points out that the very limited body of time modification research has been conducted primarily with disabled, not LEP, students).

5.2 Recommendations Concerning Greater Inclusiveness

To the end of achieving greater inclusiveness of test and questionnaire data from special needs populations, specific recommendations are made concerning the following topics:

- Accommodations in testing
- Minimization of floor effects
- Use of parallel verbal/nonverbal items in assessment
- Use of performance assessments and teacher ratings
- Test and questionnaire translation into other languages
- Use of ancillary data
- Imputation of missing test data
- Reassessment of eligibility status over time
- Sensitivity of inclusion strategies to age of the cohort
- IEP/LEP oversampling or parallel special studies
- Refining eligibility definitions and protocols
- Maintaining consistent markers across databases to facilitate comparison
- Conduct intensive substudies of excluded cases

Below, we address each of these thirteen recommendations in turn.

Accommodations in testing. Strategies that can be employed to increase the inclusiveness of direct testing programs are described in Ingels (1991); Spencer (1991); McGrew, Thurlow, Shriner, and Spiegel (1992); Ingels (1993); and McGrew, Thurlow, and Spiegel (1993). Special accommodations to facilitate participation in direct assessments include the following: extended time limits or breaking the test into split/multiple sessions; small group or one-on-one administration; translation (including the use of sign language); paraphrase, reading to the student (directions, or questions and content), using visual aids; allowing the use of dictionaries; taking dictation from the student; providing special acoustics,

furniture, visual magnification or auditory amplification devices; and providing large print or Braille versions of instruments. Some accommodations are costly; others are not. To the extent that one wants to improve the accuracy of overall estimates of the achievement or characteristics or educational inputs of the overall student population, costs can, as Spencer notes (1991), be minimized by using lower sampling rates for students who are expensive to assess in very large-scale assessments. However, for most sample surveys, and especially to the degree that disability subgroups and LEPs are of specific interest, oversampling within a general population study, or a special population study, will be the indicated approach.

While there are many means by which assessment instruments could be made accessible to the handicapped--administering verbal items by sign language for the hearing impaired, developing non-visual cognitive measures for items that require the use of visual stimuli for the blind, and so on-- there may remain some instances in which it will not be possible to develop an accommodation or equivalent item to work around the handicap. Furthermore, it is clearly possible that students who receive accommodations might, despite thorough efforts, not be taking comparable assessments. Absence of data points must be handled analytically by adjusting scores or analyzing the data separately for children with different types of disabilities. All accommodations must be pretested with substantial IEP (and LEP) field test samples.

Special note should be made of the NAEP 1995 field test, which was designed to examine the feasibility of several modifications or accommodations, and the comparability of scores: Braille, large print, one-on-one administrations, oral administration, oral answering, variable testing time, and multiple administration.

Establishing an appropriate floor for student assessments and questionnaires. Although there are many high-ability special education students, on average, in high school, students with disabilities take less of the academic curriculum and more remedial work than their typical peers[36] (as do many other students who are not in special education), and (both in elementary and high school) score lower on achievement measures. Although some LEPs are outstanding achievers, overall, limited English-language proficiency also tends to be an impediment to placement in the high school academic curriculum and to high levels of academic achievement. One implication of excluding fewer students is that the number of students scoring toward the bottom of assessment measures might increase. Likewise, more students with reading comprehension problems would be asked to complete questionnaires. Special efforts may be required to ensure that all survey forms can be comprehended by the student population, and that assessments do a good job of measuring the abilities of students who score toward the bottom of the distribution.

[36] Per Hayward and Thorne (1990, pp.ii-iii) "The enrollment patterns of students receiving special education services in high school differ from those of their nondisabled classmates. They earn considerably fewer credits in academic subjects, more credits in vocational education, and about the same in personal/other courses...a relatively high proportion (44 percent) of the academic credits they earn in regular classes are in basic or remedial courses...Given their concentration in basic/remedial courses and the finding that they earn considerably fewer credits in academic subjects, it appears that their high school academic programs differ considerably from those of their nondisabled classmates."

In general, longitudinal testing programs have made vigorous efforts to avoid "floor effects" since these are a major obstacle to the accurate measurement of achievement growth over time.[37] In NELS:88, one useful strategy for minimizing floor effects was provision in follow-up rounds of different reading and math test forms that were tailored to the prior round ability estimate.[38] This adaptive approach can be carried even further through use of more advanced technology, such as computer-assisted testing. While further work in making achievement measures more adaptive is clearly desirable, further methodological work is required as well in ensuring that questionnaires can be completed by students with poor reading comprehension. Such methodological work should be incorporated into the field test plan of new studies as they are launched.

Use of parallel verbal/nonverbal items in assessment. Particularly with very young children just entering school (who sometimes have limited exposure to a language code) and perhaps also with both younger and older students who suffer a language impairment or lack proficiency in the language of testing[39] it may be desirable to employ nonverbal (that is, purely conceptual) measures of children's knowledge of subject matter areas such as science and social studies and of quantitative achievement. [40] Indeed, verbal measures of quantitative skills may fail to fully capture the mathematical skills that children from low socioeconomic status homes bring with them to school. Jordan, Huttenlocher, & Levine (1992) show evidence that young children from impoverished backgrounds perform as well as middle-class children on nonverbal calculation problems but not on verbal calculation problems such as story problems and number fact problems. If the ability to test knowledge and skill conceptually (nonverbally) is important in the assessment of children whose knowledge may exceed their language skills (such as

---

[37] Measuring the achievement level of students at the bottom of the score distribution may often have been seen as less important for cross-sectional assessments. This is the case not only because cross-sectional indicators studies can be content with group-level trend data and do not have to measure individual-level growth in achievement over time, but also because a standards-based cross-sectional assessment may need only to be able to determine whether low-scoring students do not meet the prescribed standards, rather than determining how much subject matter knowledge they command.

[38] See Ingels, Scott, Rock, Pollack & Rasinski, 1994 (Chapter 6) or Rock & Pollack, 1995, for a description of the NELS:88 tests and their psychometric properties.

[39] Even if it is granted that the first objective of an assessment program is to ascertain students' competency in school subjects in the English language, nonverbal tests may be of assistance in helping to gain understanding of the school learning processes of bilingual and limited English proficient students. In a recent paper Hakuta (1994) reviewed basic research on second language learning and bilingualism. He notes that skills and knowledge learned in one language easily transfer to another language, e.g. the notion of velocity having been learned in Spanish can be transferred to English without having to relearn the concept. "In essence, in the act of learning concepts and skills, people form a schema that is independent of the specific language of presentation, even though the act of learning can involve active recruitment of the language to regulate thinking" (Hakuta, 1994). Thus in assessing what children know and how they may transfer this knowledge to other linguistic contexts, it is of interest to distinguish what they know from the language in which they know it.

[40] The general point here is that if having a large vocabulary is important to success in answering conventional test questions (Anderson and Freebody, 1981, report correlations between vocabulary scores and achievement and aptitude test scores from 0.71 to 0.98), then children's level of knowledge of the physical, biological and social world as well as their quantitative skills may be masked if their language competence levels are low. Because of the language-embeddedness of the school curriculum, we have stressed this problem only in connection with the assessment of some young children just entering school, for children with certain kinds of disabilities, and for students with limited English language proficiency. Nevertheless, the same point is sometimes made in the context of older youth as well. For example, R.J. Murnane and F. Levy (*Harvard Educational Review*, 63(1), spring 1993, p. 7) bring up this point in connection with the NLS-72 and HS&B tests: "We found that the scores on the mathematics and reading tests were highly correlated, in large part, because answering many questions on the mathematics test required reading skills. This high correlation made it impossible to obtain reliable estimates of the relative importance [to later employment outcomes] of reading and mathematics skills."

children with verbal learning disabilities, language minority children who lack proficiency in English, and children entering school from impoverished backgrounds), it may be desirable to develop and employ parallel verbal forms. Although teachers may sometimes adopt more conceptually-oriented instructional methods for students with limited language skills (or limited English language skills), schooling, like the larger society, is an environment that is dominantly shaped by verbal concepts. Nonverbal items therefore cannot replace verbal items and may be of decreasing relevance in later years of schooling. Using IRT methods[41] to put verbal and conceptual versions of test items on the same scale may be indicated.

The utility of a nonverbal measure can most readily be demonstrated to the extent that parallelism to a verbal measure can be maintained. This allows an observed systematic difference between the measures, such as high-SES or fully-English-proficient students performing at the same level on both measures but low-SES or LEP students performing on average better on the nonverbal measure, to be interpreted. Nonetheless, because of differences between verbal and nonverbal formats, complete parallelism may not be possible. However, it may be argued that the areas where parallelism is not possible separately represent knowledge and skills that should be assessed, and that a dual approach represents the best means to measure the full range of relevant knowledge. A brief examination of kindergarten-level mathematics knowledge may illustrate both points -- the potential for parallel assessment, and areas where parallelism cannot be achieved.

Items reflecting the early mathematics curriculum might be classified by content and by process (for example, the three content categories *whole number concepts*, *operations*, and *other math concepts* [geometry, fractions, measurement, time, graphing, probability, statistics], and the three processes, *knowledge/skill*, *understanding/comprehension*, and *problem solving/inference*). Given this matrix of contents and processes, one can examine how far verbal and nonverbal items may measure the same knowledge. For example, if whole number concepts are looked at at the process level of knowledge/skill, both verbal and nonverbal items can be used to assess this area, though in slightly different ways. Verbal items would tap mastery of the conventional number system by testing children's counting skills and ability to recognize the written symbols that represent number. Nonverbal items would capture children's grasp of the abstract ideas underlying the conventional system; this would be measured through their ability to recognize numerical equivalencies--a task requiring a number concept knowledge but not a mastery of the conventional number system. At the next process level for whole numbers, understanding/comprehension, items falling in this cell must test basic number concepts but require children to recognize numerical relationships. Both verbal and nonverbal items may do this by testing children's understanding of ordinality and ordinal numbers. Such items rely on children's understanding of number and their ability to compare and order numerical units. Considerable overlap can be achieved in verbal and nonverbal items in both content and process. Nonetheless, the selection of parallel verbal versus nonverbal items is sometimes constrained by format. For example, because the conventional system cannot be tested nonverbally, some important skills that are emphasized in the school curriculum, like counting and recognizing and writing numerals, cannot be nonverbally assessed. Conversely, because verbal items are used to apply math concepts to practical situations, they assess children's understanding of abstract mathematical ideas only in conjunction with their verbal skills. Where items cannot be made parallel, it is because the skills tested are not amenable to evaluation in both formats.

This is another area that would benefit from future methodological work. Current plans for the 1996-97 kindergarten through second grade field test of the Early Childhood Longitudinal Study include use and comparison of parallel verbal and nonverbal items in mathematics and general knowledge.

[41] See Hambleton (1989) for a discussion of Item Response Theory (IRT) and classical test models.

Use of performance assessments and teacher ratings. There are two main devices by which students who cannot be assessed directly by conventional means might be included in testing programs. One means is through alternative assessment of various kinds (for example, performance, portfolio, and extended response). Students can also be indirectly assessed, through the reports that may be collected from teachers (or parents) and classroom (or home) observers of the child. In the case of both alternative assessments and teacher ratings, the goal should be to put them on the same vertical IRT scale as the direct assessment measures. (This recommendation mirrors the approach discussed in using both verbal and "nonverbal" or conceptual measures in parallel, above).

Performance Assessment. If learning is taking place it can be measured, even if it cannot be measured through traditional multiple choice tests.[42] If one is willing to assess student portfolios or provide a range of opportunities to directly exhibit knowledge, one can be very inclusive indeed, though being so may prove expensive. An important opportunity to extend the inclusiveness of testing programs is provided by the further development and refinement of varieties of performance assessment.[43] Although performance assessment provides greater opportunities than traditional multiple choice testing for inclusion of the full range of students, for more costly performance-based measures, cost/validity ratios may sometimes be high if employed for all students. However, such measures surely could be considered for special groups that would otherwise be excluded. Alternative assessment of students who would otherwise be excluded will have maximum value if bridges can be built from these measures to the main test (that is, both the main and alternative measures can be put on the same scale). New methods that allow for vertically scaling[44] mixtures of "right answer only" (for example, from a multiple choice test) and "partial credit" (for example, from a performance assessment) scoring (Muraki, 1992) may offer greater testing flexibility in working around specific handicaps. In general, because of inter-rater reliability issues, and constraints of testing time that put a premium on broad coverage with good vertical scaling properties rather than intensive but narrow assessment exercises, performance assessments may prove more difficult to integrate into longitudinal than cross-sectional assessment programs.

Teacher Ratings. Teacher "comments" were collected in HS&B and teacher ratings of students on various dimensions were an important feature of the NELS:88 teacher questionnaire. However, teacher ratings or evaluations can be taken further, to provide systematic, scalable assessments of students' socioemotional development and of their cognitive development and subject matter achievement. Perhaps especially for young children, rating scale data might also be supplied by parents. Indirect assessment can also be provided through classroom or home observers. Our discussion will concentrate on teacher assessments of the pupil, since these have an appealingly wide range of application at an attractive monetary cost, and because the classroom constitutes so powerfully proximal and specific a learning environment for the child.

---

[42] An essay question, for example, supplies an inherently fairer test of the language skills of a dyslexic than does a multiple choice test such as was used in HS&B and NELS:88.

[43] Performance (or "alternative") assessment requires the student to execute a task rather than to choose an answer from a list of alternatives. Students could be asked to design a physics experiment or generate hypotheses, show the reasoning behind the solution to a mathematics problem, write an essay, or converse in a foreign language. Open-ended questions and extended tasks are types of performance assessment. Portfolio assessments may also be used to collect performance-based work. Portfolios present a sample of the student's performance over time.

[44] It is assumed that variables of interest within a longitudinal design will follow a developmental model and be capable of vertical scaling (within an IRT model) in such a way as to provide a metric that can be used to define quantitative indices of growth.

There are a number of formats in which teacher evaluations of students can be obtained. One is to present teachers with specific items, or tests, and ask the teacher to judge how well the student would perform on them (e.g., the teacher would predict student percentile scores). Another approach is to have teachers judge students' global performance within a subject area, given various qualitative ranking options. Questions may be asked about factors thought to influence achievement as well. In NELS:88, for example, teachers were asked to rate students' motivation, timely completion of homework assignments, performance relative to ability, attentiveness, absenteeism, and so on (on the NELS:88 ratings, see Shouse, Schneider & Plank, 1992; and Ehrenberg, Goldhaber & Brewer, 1995). Finally, teachers can be asked very specific questions about student cognitive performance, e.g., using a five-point scale to rate how well a first grader demonstrates an understanding of time, makes logical predictions when pursuing scientific investigations, uses strategies to add and subtract one and two digit numbers, or makes reasonable estimates of quantities.

Teacher judgments of student academic performance are widely used in research and have been given an increasing role in large-scale assessments (Stiggins, 1987). In a comprehensive review of teacher-based judgments of academic achievement, Hoge and Coladarci (1989) report that "On the whole the results revealed high levels of validity for the teacher-judgment measures." (Hoge and Coladarci report that teachers' judgments and tested achievement correlated as highly as .92, with a median correlation of .66; for another thorough recent review of the teacher judgment literature, see Perry & Meisels, 1996.) This conclusion is echoed in more recent work (Calfee & Hiebert, 1991; Kenny & Chekaluk, 1993) although the accuracy of teacher reports, particularly in certain contexts and for certain purposes, has also been questioned.[45]

In principle, direct measures of a child's cognitive performance and indirect measures (say from teachers) can substitute for each other since (if dimensionality requirements are satisfied) both can be put on the same IRT scale. In practice, estimating and comparing the reliability and validity of alternative measures of relevant constructs requires careful field testing and psychometric analysis.[46] A particular problem with this approach is that the direct assessment measures may be more scale invariant and less directly sensitive to differences in school-level teacher expectations, while there may be a lack of scale invariance in the teacher measures, given systematic variability in how rigorously or stringently various

[45] Hoge and Coladarci (1989), for example, having concluded that current research supports the notion that teacher-based judgments of achievement have generally high validity, note that "The studies revealed, however, some variability across teachers in accuracy levels and suggested the operation of certain other moderator variables." Of particular note are studies that question teachers' ability to discriminate between low achievement and specific learning disabilities (Hoge & Butcher, 1984; Salvesen & Undhein, 1994) although other concerns have been expressed as well, including teachers' ability to discriminate between achievement and motivation, their knowledge of the cognitive domains to be evaluated, the influence of their expectations on student outcomes, the possibility that teachers are better judges of older (and of high achieving) than of younger learners, and the influence of factors such as student behavior, socioeconomic status, gender, or race on their evaluations (see, for example, on achievement, Silverstein et al. 1983; Sharpley & Edgar 1986; Bennett, Gottesman, Rock & Cerullo 1993; on evaluations of effort, Shouse, Schneider & Plank, 1992; and on the relationship between race and gender, and overall subjective evaluations of students, Ehrenberg, Goldhaber & Brewer, 1995). The high correlation between teacher evaluations and achievement test results suggests that teacher biases are minimal (Salvesen & Undhein, 1994). Heed should be taken of Hoge and Coladarci's general point -- that although teacher evaluations are valid, they can be improved and that the most accurate information comes when highly specific (and, perhaps, criterion-referenced) information is sought.

[46] This would include examination of intercorrelations among the two modes of measurement and confirmatory factor analysis to estimate factor loadings of the two independent sources of information on a single factor. Also, alternative maximum likelihood confirmatory models that define a factor based on the direct measures and a second factor based on the indirect measures should be tested.

BEST COPY AVAILABLE

teachers across various schools, districts and regions apply student ratings. Such an attenuation in the correlation between the two measurement modes can be minimized in national assessments by having very explicit behaviors defined for each of the criterion-referenced points in the indirect measures. Moreover, as long as there are direct measures that can be related to the indirect measures and that can be scaled together, much of the variance due to local expectations can be dealt with in much the same manner that IRT has been used for adjustment in student grades (in effect, another form of teacher rating) across different schools, departments or programs (Stricker, Rock, Burton, Muraki & Jirele 1992, 1994; see also Young 1993 for a comprehensive review of the literature on grade adjustment methods). In terms of scalability of the measures, as noted in our discussion of performance assessment, theory and software have been developed to estimate item parameters using marginal maximum likelihood procedures as applied to mixtures of right-only items and items or tasks that follow multiple categories or ordered responses as in a rating scale format (see Muraki, 1992). This capacity to put right-only answers and continuous scores (such as found in free response, portfolio assessment, and teacher ratings of child performance) on the same vertical scale should be exploited.

At the same time, clearly more research is required on use of teacher measures with the special populations that are the focus of this report, since overall accuracy of teacher report does not guarantee such results will hold for various categories of students with disabilities or for students whose English language proficiency is limited. In particular, it would be useful to know more about the difference in quality of ratings provided by teachers with knowledge, training, and experience of special populations, versus teachers who lack such knowledge, training, and experience.

Test and questionnaire translation into other languages. In light of the small numbers of Spanish-language student questionnaires completed in HS&B and NELS:88, we recommend that further methodological work be done on the best method for presenting to Spanish-mother tongue students the option of completing the questionnaires in Spanish, as well as an assessment of how many language-excluded Spanish speakers unable to complete a written Spanish questionnaire could respond to an orally-administered version.

Since a major goal of the education system is to provide all students with a high level of skills in the use of English, it would seem desirable for any longitudinal study to assess English language literacy for LEPs with each round, until the student is sufficiently proficient to complete the English-language assessments.

Given the difficulty of achieving equivalent measures across languages, the large numbers of languages spoken by LEP students, the amount of instruction LEP students typically receive in English, and the fact that not all are literate in their native language and many do not receive native language instruction, translation would not appear to offer a quick and easy solution to the problem of assessing the subject matter achievement of this group. However, results of the special LEP inclusion feasibility trials for the 1995 NAEP field test should be carefully studied. The 1995 field test plan involved two LEP studies, one using a Spanish-English side-by-side assessment, the other a Spanish-only assessment at grades 4 and 8.

Use of ancillary data. Ancillary data should be collected regardless of whether the sample member can be directly assessed. These data should encompass surveys of parents, special education or bilingual education teachers, regular teachers, school and program administrators, and school records.

However, for IEP and LEP students, data should be collected over and beyond that obtained in regular parent, teacher and school questionnaires. There is a good deal of data about LEP and IEP students that should be collected, but has not been gathered in studies such as HS&B and NELS:88. For handicapped students, much more information should be gathered about the individual's physical, cognitive, and psychosocial functioning. At minimum, the data elements in the NAEP Excluded Student Questionnaire (type of disability, severity, functional grade level in math and English, percentage of the school day student is served by a special education program, and so on) would be of extraordinary value to obtain for *all IEP and LEP students, regardless of their eligibility status*. In addition, it would be desirable to obtain school reports of the sample member's behavior in educational and social settings (that is, the degree to which there are behavior problems, and whether they lead to limitations in daily activities). For LEP students, information about how long the student has been served by a special language program, how much of the student's instruction is received in the native language, and the teacher's rating of the student's proficiency along the separate dimensions of speaking, understanding, reading and writing English, should be obtained.

At the high school level, almost 90 percent of special education students receive grades that can be captured in records studies (OSEP 1994). Transcript data collections should invest in the coding of special education courses and in clearly describing the IEP of each special education student, as was done in the NAEP 1987, 1990 and 1994 High School Transcript studies.[47] There is a particular problem with the approach to collecting transcripts that was implemented in HS&B and NELS:88 that affects LEP and IEP students especially and should be avoided in future studies. Both HS&B and NELS:88 collected secondary transcripts starting in the fall following the spring term of the senior year. Consequently, records collected for students who fell behind the modal progression for their class were incomplete, that is, did not span the length of high school. Hayward and Thorne (1990) report that only 68 percent of disabled (compared to 87 percent of nondisabled) students graduate on time; LEP students also are disproportionately likely to repeat a grade (or drop out). The data in this report show that for excluded eighth grade cohort LEP and IEP students, only 62.4 percent of the excluded 1988 eighth grade cohort was in school in 1992; presumably some number of the eighth grade cohort excluded students who were 1992 dropouts will return to high school. Of this portion of the eighth grade excluded students still in school, 57.6 percent were spring 1992 high school seniors but 42.4 percent had fallen behind by a grade or more. In order to capture additional course-taking information about excluded (as well as included) IEP and LEP students who drop out but return to school or who fall out of sequence with their peers and complete high school more slowly, a further data collection should be effected two years after the modal senior year of the cohort to complete school records collection for individuals who did not graduate with their class.

It is important to collect as much longitudinal data as possible within national education studies regardless of whether a given member of the population is ever able to complete assessment instruments. Ancillary data addresses issues of process and helps to interpret cross-sectional results for special populations. Longitudinal studies in the mold of HS&B, NELS:88 and ECLS permit us to trace the school careers of students, to continue to follow them if they drop out of school, and to measure occupational and postsecondary outcomes for them as well. Moreover, longitudinal studies permit us to ascertain to what degree disabilities and lack of English language proficiencies may be transient states, to trace entry into and exit from special programs, to compare individuals who become proficient in English with those

[47] The NAEP transcript studies encompassed both NAEP assessment-included and excluded students and the 1987 study included an oversample of students with disabilities. On the NAEP transcript studies, see Thorne et al. 1989 and Legum et al. 1992. On the NAEP, HS&B and NELS:88 transcript studies, see Ingels and Taylor 1995.

who do not, and to compare the differing educational and life outcomes of individuals who share a disability status. These purposes can be achieved even when direct assessment is not possible.

Imputation of missing test data. Particularly if a good deal of ancillary data is collected that describes a sample member's relevant characteristics, it is often possible to impute missing test data. This could be done in a situation where an individual can take one portion of a test, but, for whatever reason, not another, or does not take the test at all (a good example of this is the 1992 National Adult Literacy Survey[48] [NALS], which used imputation to estimate literacy proficiencies when test data were missing). This procedure is more attractive for producing cross-sectional group estimates than for longitudinal testing. However, imputation to avoid longitudinal panel attrition (such as filling in missing data points for a student who could not participate in the baseline or who misses one of the follow-up rounds) is a very attractive option for longitudinal studies. It would be of interest, in particular, to impute base year test scores for students excluded from the base year who proved testable in later rounds, so that these cases would be usable in cohort panel analysis. More research on the most relevant background characteristics for imputation of results for special needs populations is recommended.

Reassessment of eligibility status over time. As was done in NELS:88, the eligibility status of any excluded student should be reassessed over time. This is especially important if the study intends to generate future (post-baseline) grade-representative cohorts. If all students are included in some way (for example, information is gathered about them even if they are not tested), it will nevertheless be usually desirable to reassess the ability of test-exempted students to complete assessment instruments in later years.

Sensitivity of inclusion strategies to age of the cohort. Strategies for including special needs populations such as special education or LEP students must take account of certain differences that will obtain for different age ranges. In sheer numerical terms, many more LEP students, for example, will be encountered in early grades (say kindergarten or first grade) than in (say) the last two years of high school. For students in special education, by middle or high school, students with mild mental disabilities have in large measure been identified. This is not the case in the early grades of schooling. Table 7 (below) suggests a pattern of increasing identification of learning disabled children over the first few years of school. There are several implications of this pattern for assessment in kindergarten and early grades. One is that many individuals who may face a barrier to assessment may not yet have been identified by their schools. Others may have been identified as special education students yet be misclassified within the special education categories.[49] One must be very careful to ensure that children with disabilities are not subjected to testing or assessment procedures that are inappropriate or clearly impossible for them to perform. At the same time, longitudinal assessment data will prove valuable by providing a window into the classification process itself, and will illuminate natural experiments[50] concerning the effects of identification or misidentification. Classification errors potentially have major implications for children's

[48] On NALS see Kirsch, Jungeblut, Jenkins & Kolstad (1993); for details on imputation procedures see the companion technical report.

[49] Table 7 suggests a good deal of fluidity in classifications, with speech impairment classifications declining as learning disability increases. Because special education funding and services are tied to classification, yet there is reluctance to expose children to the negative effects of labeling, speech impairment may be used as a non-stigmatizing holding category, particularly for students whose developmental delays present classificatory uncertainties. Later some of these students are reclassified as learning disabled.

[50] A longitudinal assessment can plot the educational trajectories of learning disabled students who are improperly classified, as well as of students who are misclassified as learning disabled.

later school progress and are therefore critically important to know about. It is also of interest to know how identification varies as a function of type or severity of disability and child, family, school and community characteristics.

Table 7
Public Elementary School Students in Special Education, 1991-92

Numbers Served	**6 yr olds**	**7 yr olds**	**8 yr olds**	**9 yr olds**	**10 yr olds**	**11 yr olds**
Mental Retardation	19603	26435	32880	37349	39169	42054
Speech Impairment	197984	201793	178547	137942	96839	63179
Visual Impairment	1173	1462	1563	1687	1649	1659
Emotional Disturbance	7646	13567	19996	25941	30170	34663
Orthopedic Impairment	4646	4380	4335	4099	3855	3517
Other Health-Impaired	3796	4455	4912	5070	4932	4728
Learning Disabilities	32838	84359	152012	204088	232045	244551
Deaf-Blindness	59	57	60	57	50	71
Multiple Disabilities	7324	6954	7353	7188	7106	6684
Hearing Impairment	3122	3563	3280	4102	4000	4045
Autism	484	428	383	300	250	242
Traumatic Brain Injury	11	13	12	18	16	23
	278,666	347,466	405873	427841	420081	405416

Table 7
Public Elementary School Students in Special Education, 1991-92
(Continued)

Subcategories as Percentages of Total IEP Population	**6 yr olds**	**7 yr olds**	**8 yr olds**	**9 yr olds**	**10 yr olds**	**11 yr olds**
Mental Retardation	7.0	7.6	8.1	8.7	9.3	10.3
Speech Impairment	71.0	58.0	43.9	32.2	23.0	15.6
Visual Impairment	0.4	0.4	0.4	0.4	0.4	0.4
Emotional Disturbance	2.7	3.9	4.9	6.0	71.8	8.5
Orthopedic Impairment	1.6	1.3	1.1	0.9	0.9	0.9
Other Health-Impaired	1.3	1.3	1.2	1.2	1.2	1.2
Learning Disabilities	11.8	24.3	37.5	47.7	55.2	60.3
Deaf-Blindness	0.0	0.0	0.0	0.0	0.0	0.0
Multiple Disabilities	2.6	2.0	1.8	1.7	1.7	1.6
Hearing Impairment	1.1	1.0	0.9	1.0	1.0	1.0
Autism	0.1	0.1	0.1	0.1	0.1	0.1
Traumatic Brain Injury	0.0	0.0	0.0	0.0	0.0	0.0

Percentage of the Public School Population	**6 yr olds**	**7 yr olds**	**8 yr olds**	**9 yr olds**	**10 yr olds**	**11 yr olds**
	7.9	9.9	11.6	12.2	12.0	12.0

Source: U.S. Dept. of Education, Office of Special Education Programs, Data Analysis System.

IEP/LEP oversampling or parallel special longitudinal studies. A strong argument can be made that national samples of students should include *all* students. However, in order to be able to report on rare subgroups, one must move beyond a policy of non-exclusion to one of oversampling.[51] There are two possible strategies for obtaining large numbers of LEP or special education students (including analytically robust samples of subgroups, for example, LEPs of Asian, Puerto Rican, Mexican, or other Hispanic background; students with learning disabilities, serious emotional disturbance, speech impairments, and so on). One option is oversampling within a national study such as NELS:88 or ECLS; NELS:88, for example, screened for and oversampled hearing impaired students in a special supplement conducted for Gallaudet University and gathered audiological and other records in connection with hearing impaired students. The alternative strategy is to draw special samples and conduct separate studies as was done, for example, by conducting the National Longitudinal Transition Study of Special Education Students starting in 1987. When separate studies are conducted, we recommend that they parallel national education longitudinal studies in two ways, so that comparability can be maximized: that they be roughly parallel in time, and that design linkages be pursued in the design of instruments and procedures. To the extent that any category of student is unrepresented in a NELS:88-like study, it should be possible to identify that category of student in a special study and make meaningful comparisons using the two databases. It is also possible, through nested substudies within a larger national study, to gather specialized program and other supplemental data to extend capacity to study special populations.

Refining eligibility definitions and protocols. Further work needs to be done in developing eligibility definitions. Even if one excludes no one from a study -- by virtue of including them in the collection of transcripts, teacher or parent reports, and so on -- decisions must still be made about who can be assessed. To be reliable, guidelines should be specific to the extent possible, and tied to objective performance criteria that relate to the assessment or questionnaire tasks. Attention must also be given to monitoring the implementation of guidelines so that corrective action can be taken. In addition to guidelines for inclusion, guidelines for accommodations must also be developed.

The 1995 NAEP field test experimented with two different approaches to inclusion/exclusion criteria. In the mathematics sample, all students with disabilities were to be included, unless no appropriate accommodation was feasible. In the science sample, students with disabilities were to be included based on a set of curriculum-matching criteria. Results obtained from these two approaches should be studied carefully.

Maintaining consistent markers across databases to facilitate comparison. There are certain benefits to developing consistent definitions of eligibility across national databases. There are also benefits to maintaining compatible classificatory schemes. One benefit of common eligibility definitions is increased capacity for test equating. An example of such equating is provided by the NELS:88 second follow-up. Both NELS:88 and NAEP assessed high school seniors in the spring of 1992. A NAEP-

[51] About 12 percent of the public school population and 9-10 percent of the total school population receives special education services. A NELS:88-size sample (the base year was 26,432 students in 1,052 schools, of whom 24,599 students participated) would yield around 2,643 special education students (26,432 x .10) and around 2,460 completed cases. Just over half of these sample members would be learning disabled students (per OSEP 1994, Table 1.4, for the 1992-93 school year), followed by individuals with speech impairments. The completed case yield for learning disabled students would be 1,257, with 531 participants for the speech-impaired category. Yields for the mentally retarded would be 282 completed cases, for serious emotional disturbance 214. For the rare disability populations yields would be yet lower. For orthopedic impairments the yield would be 27, for visual impairments 12. Other populations are much rarer still (autism, deaf-blind, traumatic brain injury). Numbers for any disability subcategory could be increased by oversampling, that is, students within a disability category within any sampled school could be taken at a higher rate, up to 100 percent.

equated mathematics score (calibrated on the NELS:88 sample of seniors) was reported for every NELS:88 sample member who completed a 1992 mathematics test. (A benefit of putting NELS:88 results on the NAEP scale is that it permits the NAEP results to be connected to the rich school and home context and process variables in NELS:88.) A further benefit of implementing the NAEP scale in NELS:88 is that it adds a dimension of interpretability to NELS:88, given that the NAEP scales are so well known to the public and educational community as a measure of what students can and cannot do.) Another example of such equating is provided by the NELS:88 first follow-up (Rasinski, Ingels, Rock & Pollack, 1993), which put High School and Beyond (HS&B) 1980 sophomore and NELS:88 1990 sophomore mathematics results on the same scale so that trend comparisons could be drawn. The equating of two measures can be carried out if either or both of two conditions hold: the different forms are administered to equivalent samples, or there are items common to both forms. Ideally, one will have equivalent samples and (at least) some small set of common items as an additional check on the acceptability of the equating results. While some differences can be treated analytically, it is highly desirable that the samples be as similar as possible. This argues for likeness of inclusion and exclusion rules whenever possible. Of course, some additional flexibility comes from the capacity to flag special sample groups and combine and recombine them for purposes of specific comparisons. But here too attention to comparability in definitions across national databases is of the essence, since such mappings also depend on use of compatible (though not necessarily identical) categories. It would be particularly beneficial if longitudinal school studies, and school to work transitions studies, used consistent markers for students with disabilities, and asked some comparable questions. Consistency should be achieved not just within U.S. Department of Education studies, but also with studies such as the Department of Labor's National Longitudinal Survey, Youth Cohort, which anticipates launching a new NLSY cohort in 1997.

There are 13 Federal disability categories (two rare but distinct disabilities, autism and traumatic brain injury, were added in 1991 and do not appear in earlier statistical reports). Students who receive special education services can be categorized in terms of these distinctions since they are contained within the IEP.[52] Nevertheless, the National Center for Educational Outcomes (NCEO) found that there was considerable variation in use of disability categories across the 19 national data collection programs NCEO studied (see OSEP 1994, Table 5.8). Such variability creates a problem both of loss of information in some data collections, and lack of comparability, across others. Greater consistency in the use of standard categories would be beneficial.

While for most purposes, special needs populations can be defined in terms of those individuals receiving specific services such as special education, bilingual education, and so on, it is also of value to be able to define these populations analytically. To an important extent, a policy of maximum inclusiveness provides a check on the accuracy of the classifications in use in the schools. Thus while local criteria for who receives bilingual education services or may be enrolled in English as a second language may vary, a national study can provide an objective measure of English language proficiency. Likewise, there may be some amount of misclassification of special education students, especially those with learning disabilities. For young children in particular it may be difficult to distinguish slow maturation from learning problems. Diagnoses of learning disability are subject to (to borrow the language of statistical hypothesis testing) both Type I and Type II errors. Children who later turn out to have learning disabilities are often undetected at early testing, and some who are diagnosed as having learning disabilities, later turn out to have been slow maturers. Achievement test data may permit a

[52] One of the following thirteen categories will be listed on each IEP as the student's primary disability: learning disability; mental retardation; speech or language impairment; serious emotional disturbance; orthopedic impairment; deaf; hard of hearing; visual handicap; deaf/blind; other health impairment; multiple handicap; traumatic brain injury; autism.

second reading on which students do or do not have specific learning disabilities, regardless of how schools have classified them. The preferred analytic strategy for a longitudinal study with a test component -- using individual growth curves[53] -- focuses attention on growth rates over the time span of the study and permits analysts to separate out rate of growth from the level of performance. The curve for learning disabled children should in time assume a distinctive shape.

Intensive substudies of excluded cases. Finally, excluded students may be intensively studied to learn more about why they are excluded, as well as how they (and their educational outcomes) differ from students who will be assessed. Either a subsample of excluded students can be drawn for special study, or (see recommendation on ancillary data above) students excluded from assessment can be retained in other components of the study (student questionnaire, even if this means a shift from self-administered to interviewer-administered mode; contextual components, such as teacher and parent and principal reports; school records). An example of such a substudy is the recent NAEP state assessment IEP exclusion study taking place with fourth graders in four states, which will report on how included students differ from excluded, based on results from the Woodcock-Johnson (see Glossary).

5.3 Recommendations Concerning Minimizing Other Sources of Undercoverage

Recommendations with respect to baseline transfer students. While the largest source of undercoverage in the NELS:88 base year is the five percent of the sample that was excluded from participation, an additional source of undercoverage is the under-representation of transfer students. An expensive and logistically complicating solution to the problem of transfers between the time of sampling (in HS&B and NELS:88, the fall term) and surveying (in HS&B and NELS:88, the spring term) of students would be to follow all students once they have been selected.

A more cost-efficient strategy, and one which maintains design simplicity, is to continue the strategy of excluding transfers-out and sampling from the transfers-in in a sample update just prior to the school's survey session, but to accommodate undercoverage of transfer students in the weighting. One should collect race/ethnicity, gender, and other basic information about the sample at the time of initial sampling. Weights for transfer-out students should be calculated. The estimated undercoverage of transfer-ins would be accounted for by modifying the weights of this group appropriately. Indeed, we recommend that information on a range of key characteristics be collected on all selected students from school records at the time of sample selection not only to accommodate the problem of asymmetric capture of transfers out and in, but also to assess and adjust for the impact of nonresponse. Logistic regression can be used to model the likelihood that a given student will complete the survey. With richer information on characteristics, gathered at the time of sample selection, response propensities can be used to develop adjustments that compensate for the effects of sample attrition.

5.4 Recommendations with Respect to Sampling Lists

School Frames. LEP students do not normally attend special schools that might be excluded from the sampling frame, although between six and eight percent of students with disabilities attend special schools that exclusively serve the disability population. Some 63 percent of deaf students, 35 percent of the visually impaired, and 41 percent of students with multiple impairments attended special secondary

[53] See Rogosa & Willett (1985), Ware (1985), and Bryk & Raudenbush (1987).

schools (Wagner, Blackorby, Cameto, & Newman, 1993, Table 4-1, p.4-4). Accordingly, to the extent that these special subpopulations are of independent interest, frame exclusions are a matter of major concern. However, there is a difference between giving rare populations a chance of selection into a sample, and representing them at a level that supports separate estimates; the latter goal would require oversampling. Bureau of Indian Affairs (BIA) schools, which contain eight to ten percent of the American Indian student population as well as disproportionate numbers of LEPs and special education students, were excluded from the HS&B and NELS:88 sample frames. There seems no justification for doing so; BIA schools should be included in future studies.

Student rosters. LEP students are more likely to be migrants, and more likely to go back and forth across international borders. A small proportion of special education students are sometimes in ungraded programs, hence missed in studies that sample grade cohorts. To the extent that the target population is a grade -defined cohort, there is justification for ignoring ungraded programs.

Migrant students are not a large factor in national estimates (less than one percent of the school-age total) but of greater weight in LEP estimation and of policy interest in their own right. At the very least, they should be identified at the time of sampling and sample attrition for this group studied and its impact on LEP estimates taken into account. Migrants may be oversampled within a national study although there are obstacles to cost-efficiently doing so within a NELS-HS&B type design because of the rarity of the population.

Longitudinal studies could encompass a supplemental age cohort to complement the grade cohort; NAEP has an age cohort trend sample as well as grade level samples. Age cohorts could include ungraded school populations. Ungraded students are a very rare population comprising mostly students with severe cognitive deficits. It is not clear that curriculum-sensitive tests and curriculum-relevant questionnaires truly capture the educational situation and progress of this group. LEPs (especially new arrivals) are sometimes placed in ungraded programs and some other groups of students are in ungraded programs that reflect a philosophical orientation hostile to strict grade differentiation; such students should be part of the ideal target population.

Chapter 6: Bibliography

Algozzine, B. and Korinek, L. 1985. "Where is Special Education for Students with High Prevalence Handicaps Going?" *Exceptional Children*, 51, 388-94.

Allen, T.E. (Ed.) 1989. "Handicapped Youth and the Outcomes of Their Schooling: Recent Research Advances in the United States." *International Journal of Educational Research*, 13.

Anderson, R.C., and Freebody, P. 1981. "Vocabulary Knowledge." In J.T. Guthrie, ed., *Comprehension and Teaching: Research Reviews*. Newark, Delaware: International Reading Association.

Bennett, R.E., and Ragosta, M. 1984. *A Research Context for Studying Admissions Tests and Handicapped Populations*. Princeton, N.J.: Educational Testing Service.

Bennett, R.E., and Ragosta, M. 1988. "Handicapped People." In Willingham, et al. (see below)

Bennett, R.E., Gottesman, R.L., Rock, D.A., and Cerullo, F. 1993. "Influence of Behavior Perceptions and Gender on Teachers' Judgments of Students' Academic Skill." *Journal of Educational Psychology*, 85(2).

Bruininks, R., Thurlow, M.L., and Ysseldyke, J.E. 1992. "Assessing the Right Outcomes: Prospects for Improving Education for Youth with Disabilities." *Education and Training in Mental Retardation*, 27.

Bryk, A.S., and Raudenbush, S.W. 1987. "Application of Hierarchical Linear Models to Assessing Change." *Psychological Bulletin*. 101(1).

Bürkheimer, G., et al. 1990. *National Longitudinal Evaluation of the Effectiveness of Services for Language-Minority Limited English Proficient Students*. Report to the Office of Bilingual Education and Minority Languages Affairs (OBEMLA), U.S. Department of Education. Research Triangle Park, NC: Research Triangle Institute.

Calfee, R., and Hiebert, E. 1991. "Teacher Assessment of Achievement." *Advances in Program Evaluation*, vol. 1A. JAI.

Chalfant, J.C. 1989. "Learning Disabilities: Policy Issues and Promising Approaches." *American Psychologist*.

Clay, M.M. 1990. "Learning to Be Learning Disabled." *ERS Research Digest*, pp.11-16. Arlington, VA: Educational Research Service.

CCSSO (Council of Chief State School Officers). 1990. *School Success for Limited English Proficient Students: The Challenge and State Response*. Washington, D.C.: Author.

CCSSO (Council of Chief State School Officers). 1992. *Recommendations for Improving the Assessment and Monitoring of Students with Limited English Proficiency*. Washington, D.C.: Author.

Ehrenberg, R.G., Goldhaber, D.G., and Brewer, D.J. 1995. "Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Education Longitudinal Study of 1988." *Industrial and Labor Relations Review*, *48* (3), pp. 547-61.

Fedoruk, G.M., and Norman, C.A. 1991. "Kindergarten Screening Predictive Inaccuracy." *Exceptional Children*, 57.

Fleischman, H.L., and Hopstock, P.J. 1993. *Descriptive Study of Services to Limited English Proficient Students*. Prepared for Office of the Under Secretary U.S. Department of Education by Development Associates, Inc., Arlington, VA.

Gandara, P., and Merino, B. 1993. "Measuring the Outcomes of LEP Programs: Test Scores, Exit Rates, and Other Mythological Data." *Educational Evaluation and Policy Analysis*, 15(3).

GAO (U.S. General Accounting Office). 1993. *Student Testing: Current Extent and Expenditures, With Cost Estimates for a National Examination*. Washington, D.C.: GAO/PEMD-93-8.

Green, P.J., and Scott, L.A. 1995. *"At-Risk" Eighth Graders Four Years Later*. Statistics in Brief series. Washington, D.C.: National Center for Education Statistics.

Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

Hafner, A.L. 1994. *Making Our Assessments Comprehensible to English Language Learners*. Los Angeles: California State University.

Hafner, A.L. 1995. *Developing and Modifying Statewide Assessments for LEP Students*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hakuta, K., and Garcia, E.E. 1989. "Bilingualism and Education." *American Psychologist*, 44.

Hakuta, K. 1994. Perspectives from basic research with specific reference to the inclusion of LEP students in NAEP. *Conference on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment on Educational Progress*, Dec. 5-6, 1994, Washington, D.C.

Hambleton, R.K. 1989. "Principles and Selected Applications of Item Response Theory." In R.L. Linn (ed.) *Educational Measurement*. 3rd ed. New York: MacMillan.

Hambleton, R.K. 1993. "Translating Achievement Tests for Use in Cross-National Studies." *European Journal of Psychological Assessment*, 9(1).

Hayward, B.J., and Wirt, J.G. 1989, August. *Handicapped and Disadvantaged Students: Access to Quality Vocational Education* (v.5, National Assessment of Vocational Education Final Report.) Washington, D.C.: U.S. Department of Education.

Hayward, B.J., and Thorne, J.M. 1990, October. *The Educational Programs of High School Special Education Students*. Contractor report to the Office of Special Education Programs, U.S. Department of Education. Research Triangle Park, N.C.: RTI.

Henderson, A., Daft, J., and Gutmann, B. 1989. *A Summary of State Chapter 1 Migrant Education Program Participation and Achievement Information, 1986-87.* Washington, D.C.: Decision Resources Corporation.

Henderson, A., Abbott, C., and Strang, W. 1993. *Summary of the Bilingual Education State Agency Program Survey of States' Limited English Proficient Persons and Available Educational Services 1991-1992.* Report of the Special Issues Analysis Center to the Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education.

Hoachlander, E.G. 1992. *Participation in Secondary Vocational Education, 1982-87.* Washington, D.C.: National Center for Education Statistics.

Hodapp, R.M., and Kraser, D.V. 1995. "Reflections on Using Large, National Data Bases in Special Education Research." *Exceptionality*, 5(2), pp. 103-108.

Hoge, R.D., and Coladarci, T. 1989. "Teacher-Based Judgments of Academic Achievement: A Review of Literature." *Review of Educational Research*, 59(3).

Hoge, R.D., and Butcher, R. 1984. "Analysis of Teacher Judgments of Pupil Achievement Levels." *Journal of Educational Psychology*, 76.

Houser, J. 1995. *Assessing Students with Disabilities and Limited English Proficiency.* Washington, D.C.: National Center for Education Statistics, NCES Working Paper Series.

Ingels, S.J., Rizzo, L., and Rasinski, K.A. 1989, March. *School, Individual and Item Nonresponse in the NELS:88 Base Year Survey.* Paper presented at the American Educational Research Association annual meeting, San Francisco, California. ERIC ED 312-311.

Ingels, S. J. 1991, April. *The Problem of Excluded Baseline Students in a School-Based Longitudinal Study.* Paper Presented at the American Educational Research Association annual meeting, Chicago. ERIC ED 334-273.

Ingels, S.J., and Scott, L.A. 1993, April. *Exclusion of Students with Barriers to Participation in NELS:88--Baseline Excluded Students Two and Four Years Later.* Paper presented at the American Educational Research Association annual meeting, Atlanta, Georgia, Session 25.60. ERIC ED 360-371.

Ingels, S.J. 1993, June. *Strategies for Including All Students in National and State Assessments: Lessons From a National Longitudinal Study.* Paper presented in Session 38 of the National Conference on Large-Scale Assessment, Albuquerque, New Mexico. ERIC ED 363- 645.

Ingels, S.J., Scott, L.A., Rock, D.A., Pollack, J.M., and Rasinski, K.A. 1994. *NELS:88 First Follow-Up Final Technical Report.* Washington, D.C.: National Center for Education Statistics. NCES 94-632 .

Ingels, S.J., Dowd, K.L., Baldridge, J., Stipe, J., Bartot, V.H., and Frankel, M.R. 1994. *NELS:88 Second Follow-Up: Student Component Data File User's Manual.* Washington, D.C.: National Center for Education Statistics, NCES 94-374.

Ingels, S.J., and Taylor, J.R. 1995. *Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data.* Washington, D.C.: National Center for Education Statistics, NCES Working Paper Series..

Johnson, E.G., and Zwick, R. 1990. *The NAEP 1988 Technical Report.* Report to NCES prepared by the Educational Testing Service, Princeton, N.J.

Jordan, N.C., Huttenlocher, J, and Levine, S.C. 1992. "Differential Calculation Abilities in Young Children from Middle and Low-Income Families." *Developmental Psychology*, 28, 644-653.

Kaufman, S. 1991, May. *1988 Schools and Staffing Survey Sample Design and Estimation.* Washington, D.C.: National Center for Education Statistics. NCES 91-127.

Kenny, D.T., and Chekaluk, E. 1993. "Early Reading Performance: A Comparison of Teacher-Based and Test-Based Assessments." *Journal of Learning Disabilities*, 26.

Kirsch, I.S., Jungeblut, A., Jenkins, L., and Kolstad, A. 1993. *Adult Literacy in America.* Washington, D.C.: National Center for Education Statistics.

Krashen, S.D. 1991. "Sheltered Subject Matter Teaching." *Cross Currents, 18*, 183-188.

Kronick, D. 1988. *New Approaches to Learning Disabilities: Cognitive, Metacognitive, and Holistic.* Philadelphia: Grune and Stratton.

Legum, S., Caldwell, N., Goksel, H., Haynes, J., Hynson, C., Rust, K., and Blecher, N. 1992. *The 1990 High School Transcript Study Technical Report.* Report to the National Center for Education Statistics. Rockville, MD: Westat, Inc.

Levine, S.C., Jordan, N., and Huttenlocher, J. 1992. "Development of Calculation Abilities in Young Children." *Journal of Experimental Child Psychology*, 53. 72-93.

McGill-Franzen, A., and Allington, R.L. 1993. "Flunk'em or Get Them Classified: The Contamination of Primary Grade Accountability Data." *Educational Researcher*, 22(1).

McGrew, K.S., Thurlow, M.L., Shriner, J.G. and Spiegel, A.N. 1992. *Inclusion of Students with Disabilities in National and State Data Collection Programs.* Minneapolis: National Center on Educational Outcomes.

McGrew, K.S., Algozzine, B. Ysseldyke, J., Thurlow, M., and Spiegel, A.N. 1993. *The Identification of People With Disabilities in National Databases.* Minneapolis: University of Minnesota, National Center on Educational Outcomes.

McGrew, K.S., Thurlow, M.L., and Spiegel, A.N. 1993. "An Investigation of the Exclusion of Students With Disabilities in National Data Collection Programs." *Educational Evaluation and Policy Analysis*. 15(3).

McLaughlin, M.J., and Warren, S.H. 1994. *Performance Assessment and Students with Disabilities: Usage in Outcomes-Based Accountability Systems*. Reston, VA: Council for Exceptional Children.

McMillen, M.M., Kaufman, P., Hausken, E.G., and Bradby, D. 1993. *Dropout Rates in the United States: 1992*. Washington, D.C.: National Center for Education Statistics. NCES 93-464.

Mercado, C., and Rombero, M. 1993. "Assessment of Students in Bilingual Education." In M.B. Arias and U. Casanova, eds., *Bilingual Education: Politics, Practice, Research*. Chicago: University of Chicago Press.

Mercer, C.D. 1990. "Learning Disabilities: Definitions and Criteria Used by State Education Departments." *Learning Disability Quarterly*.

Moss, M., and Puma, M. 1995. *Prospects: The Congressionally-Mandated Study of Educational Growth and Opportunity--Interim Report: Language Minority and Limited English Proficient Students*. Cambridge, MA: Abt Associates (for U.S. Department of Education).

Mullis, I.V.S., Dossey, J.A., Owen, E.H., and Phillips, G.W. 1993. *NAEP 1992 Mathematics Report Card for the Nation and the States*. Princeton, N.J., Educational Testing Service, Report No. 23-ST02, for the National Center for Education Statistics, Washington, D.C.

Muraki, E. 1992. "A Generalized Partial Credit Model: Application of an EM algorithm." *Applied Psychological Measurement*, 16.

Murnane, R.L., and Levy, F. 1993. "Why Today's High-School-Educated Males Earn Less than Their Fathers Did." *Harvard Educational Review*, v. 63.

National Council on Disability. 1993. *Serving the Nation's Students With Disabilities: Progress and Prospects*. Washington, D.C.: Author.

Office for Civil Rights (OCR). 1993. *The Fall 1990 Elementary and Secondary School Civil Rights Survey*. Washington, D.C.: Author.

Office of Special Education Programs (OSEP), U.S. Department of Education. 1994. *16th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Washington, D.C.: Author.

Owings, J.A., and Stocking, C. 1985. *Characteristics of High School Students Who Identify Themselves as Handicapped*. Washington, D.C.: National Center for Education Statistics.

Pelavin Associates. 1988. *Selection Procedures for Identifying Students in Need of Language Services*. Report to the Office of Bilingual Education and Minority Languages Affairs (OBEMLA). Washington, D.C.: Author.

Perry, N.E., and Meisels, S.J. 1996. *Teachers' Judgments of Students' Academic Performance*. Washington, D.C., National Center for Education Statistics, NCES Working Paper Series.

Pollack, J.M., and Rock, D.A. 1996. *Constructed Response Tests in NELS:88.* Washington, D.C.: National Center for Education Statistics.

Rasinski, K.A., Ingels, S.J., Rock, D.A., and Pollack, J.M. 1993. *America's High School Sophomores: A Ten Year Comparison.* Washington, D.C.: National Center for Education Statistics, NCES 93-087.

Reynolds, M.C. 1990. "Noncategorical Special Education." In Wang, M.C., Reynolds, M.C., and Walberg, H.J. (eds.), *Special Education: Research and Practice--Synthesis of Findings.* New York: Pergamon Press.

Rock, D.A., and Pollack, J.M. 1995. *NELS:88 Base Year Through Second Follow-Up Psychometric Report.* Washington, D.C.: National Center for Education Statistics. NCES 95-382.

Rogosa, D., and Willett, J.B. 1985. "Understanding Correlates of Change by Modeling Individual Differences in Growth." *Psychometrika*, 50.

Salvesen, K.A., and Undhein, J.O. 1994. "Screening for Learning Disabilities." *Journal of Learning Disabilities*, 27.

Secada, W.G. 1994. "Issues in the Development of Spanish-Language Versions of the National Assessment of Educational Progress." In *The Trial State Assessment, Prospects and Realities: Background Studies.* G. Bohrnstedt, ed. Palo Alto: Armadillo.

Sharpley, C.F., and Edgar, E. 1986. "Teachers' Ratings vs. Standardized Tests: An Empirical Investigation of Agreement Between Two Indices of Achievement." *Psychology in the Schools*, 23.

Shouse, R., Schneider, B., and Plank, S. 1992. "Teacher Assessments of Student Effort: Effects of Student Characteristics and School Type." *Educational Policy*, 6(3).

Shepard, L.A., Smith, M., and Vojir, C. 1983. "Characteristics of Pupils Identified as Learning Disabled." *American Educational Research Journal*, 20.

Silverstein, A.B., Brownlee, L., Legutki, G., and MacMillan, D.L. 1983. "Convergent and Discriminant Validation of Two Methods of Assessing Three Academic Traits." *Journal of Special Education*, 17.

Speece, D.L., and Cooper, D.H. 1990. "Ontogeny of School Failure: Classification of First-Grade Children." *American Educational Research Journal*, 27.

Spencer, B.D., Frankel, M.R., Ingels, S.J., Rasinski, K.A., and Tourangeau, R. 1990. *NELS:88 Base Year Sample Design Report.* Washington, D.C.: National Center for Education Statistics. NCES 90-463.

Spencer, B.D. 1991, November. "Eligibility/Exclusion Issues in the 1990 Trial State Assessment." In G. Bohrnstedt et al., *Assessing Student Achievement in the States: Background Studies for the*

Evaluation of the NAEP Trial State Assessment. National Academy of Education. Stanford, California: Stanford University.

Spencer, B.D. 1994. "A Study of Eligibility, Exclusions and Sampling: 1992 Trial State Assessment. In *The Trial State Assessment, Prospects and Realities: Background Studies.* G. Bohrnsted, ed. Palo Alto: Armadillo.

Stiggins, R.J. 1987. "Design and Development of Performance Assessments." *Educational Measurement: Issues and Practices.*

Stricker, L.J., Rock, D.A., Burton, N.W., Muraki, E. & Jirele, T. 1992. *Adjusting College Grade-Point Average for Variations in Grading Standards.* (Research Report 92-65). Princeton, NJ: Educational Testing Service.

Stricker, L.J., Rock, D.A., and Burton, N.W. 1993. "Sex Differences in Predictions of College Grades from Scholastic Aptitude Test scores." *Journal of Educational Psychology*, 85.

Stricker, L.J., Rock, D.A., Burton, N.W., Muraki, E. and Jirele, T. 1994. "Adjusting College Grade-Point Average Criteria for Variations in Grading Standards: A Comparison of Methods." *Journal of Applied Psychology*, 79(2).

Thorne, J., Rust, K., Burke, J., Marhsall, R., Caldwell, N. Sickles, D., Ha, P. and Hayward, B. 1989. *1987 High School Transcript Study Technical Report.* Report to the National Center for Education Statistics. Rockville, MD: Westat, Inc.

Thurlow, M.L., Ysseldyke, J.E., and Silverstein, B. 1993. *Testing Accommodations for Students With Disabilities: A Review of the Literature.* Minneapolis: National Center on Educational Outcomes, University of Minnesota.

U.S. Bureau of the Census. 1993. *Language Spoken at Home and Ability to Speak English for United States, U.S. Regions and States: 1990.* Washington, D.C.: Author, CPH-L-133.

U.S. Department of Education, National Center for Education Statistics. *The Digest of Education Statistics, 1992.* Washington, D.C.: Author.

U.S. Department of Education, National Center for Education Statistics. *The Condition of Education 1994.* Washington, D.C.: Author.

van Ijzendoor, W.J.E., and Bus, A.G. 1993. "How Valid Are Experts' Prognoses on Children With Learning Problems?" *Journal of School Psychology*, 31, 317-325.

Wagner, M., Newman, L., and Shaver, D. 1989. *The National Longitudinal Transition Study of Special Education Students: Report on Procedures for the First Wave of Data Collection (1987).* Menlo Park, CA: SRI.

Wagner, M., Blackorby, J., Cameto, R. and Newman, L. 1993. *What Makes a Difference? Influences on Postschool Outcomes of Youth With Disabilities.* The Third Comprehensive Report from the

BEST COPY AVAILABLE

National Longitudinal Transition Study of Special Education Students; prepared for the Office of Special Education Programs. Menlo Park, CA: SRI.

Ware, J.H. 1985. "Linear Models for the Analysis of Longitudinal Studies." *American Statistician*, 39(2).

Will, M.C. 1986. "Educating Children With Learning Problems." *Exceptional Children*, 52, 411-415.

Willingham, W.W., Ragosta, M., Bennett, R.E., Braun, H., Rock, D.A., and Powers, D.E. 1988. *Testing Handicapped People*. Needham Heights, MA: Allyn and Bacon.

Wolman, C., Bruininks, R.H., and Thurlow, M.L. 1989. "Dropouts and Dropout Programs: Implications for Special Education." *Remedial and Special Education*, 10.

Ysseldyke, J.E., Algozzine, B. Shinn, M.R., and McGue, M. 1982. "Similarities and Differences Between Low Achievers and Students Classified as Learning Disabled." *Journal of Special Education*, 16, 73-85.

Ysseldyke, J.E., and Thurlow, M.L., eds., 1993. *Views on Inclusion and Testing Accommodations for Students with Disabilities*. Minneapolis: National Center on Educational Outcomes.

Ysseldyke, J.E., Thurlow, M., McGrew, K., and Vanderwood, M. 1994. *Making Decisions About the Inclusion of Students with Disabilities in Large-Scale Assessments*. Minneapolis: National Center on Educational Outcomes.

Young, J.W. 1993. "Grade Adjustment Methods." *Review of Educational Research*, 63(2).1

Appendix A:

Supplemental Tables

Table A-1
Description of Characteristics of Excluded Students

	Summary				Breakdown For Expanded Sample							
	Expanded Sample		All BYIs		BYI/F2E		BYI/F2I		UNLOC		OTHER F2Es	
Characteristics	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**
Total	20070	100.0	584	4.7	334	2.5	186	1.7	64	0.5	19486	95.3
Region												
Northeast	3848	19.6	114	22.0	72	0.6	26	0.3	16	0.1	3734	18.6
Midwest	5125	25.4	119	23.2	54	0.5	51	0.5	14	0.1	5006	24.3
South	6993	35.3	173	32.7	88	0.7	70	0.7	15	0.1	6820	33.8
West	4102	19.6	178	22.1	120	0.7	39	0.3	19	0.1	3926	18.6
Race												
Asian	1305	3.5	84	7.3	61	0.2	16	0.1	7	0.0	1221	3.2
Black	2663	13.8	151	15.7	103	0.4	23	0.3	25	0.1	2512	13.1
Hispanic	2083	10.9	76	18.9	41	0.6	28	0.1	7	0.2	2007	10.0
White	13531	68.6	212	47.4	109	1.1	89	1.0	14	0.1	13319	66.4
American Indian	257	1.6	9	2.0	8	0.1	1	0.0	0	-	248	1.5
Unknown	231	1.6	52	8.6	12	0.1	29	0.2	11	0.1	179	1.2
Gender												
Male	10115	50.8	362	63.5	192	1.5	130	1.2	40	0.3	9753	47.8
Female	9955	49.2	222	36.5	142	1.0	56	0.5	24	0.3	9733	47.5
Race by Gender												
Asian-Female	628	1.7	32	3.0	25	0.2	4	0.0	3	0.0	596	3.1
Asian-Male	677	1.9	52	4.4	36	0.3	12	0.1	4	0.0	625	3.3
Black-Female	1039	6.7	24	5.3	13	0.2	9	0.2	2	0.1	1015	13.1
Black-Male	1044	7.1	52	10.5	28	0.5	19	0.4	5	0.1	992	13.0

Table A-1
Description of Characteristics of Excluded Students (cont'd)

	Summary				Breakdown For Expanded Sample							
	Expanded Sample		All BYIs		BYI/F2E		BYI/F2I		UNLOC		OTHER F2Es	
Characteristics	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**	**N**	**Wgt%**
Race by Gender (cont'd)												
Hispanic-Female	1339	5.6	75	9.1	57	0.7	8	0.1	10	0.1	1264	10.5
Hispanic-Male	1324	5.2	76	9.8	46	0.5	15	0.2	15	0.2	1248	9.4
White-Female	6716	33.8	70	15.6	37	0.8	28	0.6	5	0.1	6646	67.1
White-Male	6815	35.8	142	31.8	72	1.4	61	1.3	9	0.2	6673	65.7
Am. Indian-Female	131	0.8	2	0.4	2	0.0	0	-	0	-	129	1.6
Am. Indian-Male	126	0.8	7	1.6	6	0.1	1	0.0	0	-	119	1.3
Unknown-Female	102	0.7	19	3.2	8	0.1	7	0.1	4	0.1	83	1.1
Unknown-Male	129	0.9	33	5.4	4	0.1	22	0.4	7	0.1	96	1.3
1992 Enrollment												
In School	17299	83.3	372	62.4	232	1.7	98	0.9	42	0.3	16927	80.3
Dropout	1915	11.6	167	30.0	70	0.6	82	0.7	15	0.1	1748	10.2
Alt Completer	856	5.2	45	7.6	32	0.2	6	0.1	7	0.0	811	4.8
Grade Progression												
In Sequence	16816	80.0	345	57.6	211	1.5	96	0.9	38	0.3	16471	77.3
Out of Sequence	3254	20.0	239	42.4	123	1.0	90	0.8	26	0.2	3015	18.0
Handicap Barrier												
Physical	23		23	6.0	13	0.1	9	0.1	1	0.0	0	-
Mental	330	3.1	330	65.8	166	1.5	140	1.4	24	0.2	0	-
Linguistic	176	0.9	176	20.3	125	0.7	22	0.1	29	0.2	0	-
Unknown	55	0.4	55	8.0	30	0.2	15	0.1	10	0.1	0	-

Key: F2E = eligible in second follow-up; F2I = ineligible in second follow-up; "unloc" = unlocatable, status not ascertained.

Table A-2
1988 Exclusion Reason by Race, Sex, Race and Sex, and 1992 Enrollment

	Race/Ethnicity Unwgt N (Wgt%)					
	Asian	Hispanic	Black	White	American Indian	Unknown
Eligibility/Exclusion Reason						
Mental Exclusion						
Total	4 (0.6)	47 (8.3)	63 (19.4)	175 (59.8)	6 (2.1)	35 (9.7)
F2 Eligibility	3 (0.4)	29 (5.1)	33 (10.0)	88 (29.8)	6 (2.1)	7 (2.0)
F2 Ineligibility	1 (0.3)	14 (2.5)	26 (8.4)	75 (26.1)	0	24 (6.9)
Physical Exclusion						
Total	1 (1.4)	0	4 (25.8)	16 (64.2)	0	2 (8.5)
F2 Eligibility	1 (1.4)	0	3 (13.6)	8 (25.7)	0	1 (5.1)
F2 Ineligibility	0	0	0	8 (38.5)	0	1 (3.4)
Language Exclusion						
Total	75 (31.8)	87 (57.8)	1 (0.9)	4 (3.6)	1 (1.1)	8 (5.0)
F2 Eligibility	54 (23.1)	62 (40.4)	1 (0.9)	4 (3.6)	--	4 (2.2)
F2 Ineligibility	14 (6.5)	6 (3.5)	--	--	1 (1.1)	1 (0.3)

Table A-2
1988 Exclusion Reason by Race, Sex, Race and Sex, and 1992 Enrollment
(cont'd)

	Gender Unwgt N (Wgt %)	
Eligibility/Exclusion Reason	**Male**	**Female**
Mental Exclusion		
Total	225 (69.3)	105 (30.7)
F2 Eligibility	109 (33.2)	57 (16.3)
F2 Ineligibility	98 (31.6)	42 (12.5)
Physical Exclusion		
Total	13 (50.0)	10 (50.0)
F2 Eligibility	8 (28.6)	5 (17.3)
F2 Ineligibility	5 (21.4)	4 (20.6)
Language Exclusion		
Total	92 (51.0)	84 (49.0)
F2 Eligibility	60 (31.6)	65 (38.6)
F2 Ineligibility	17 (9.2)	5 (2.2)

Table A-2
1988 Exclusion Reason by Race, Sex, Race and Sex, and 1992 Enrollment
(cont'd)

	1992 Enrollment Unwgt N (Wgt %)		
Eligibility/Exclusion Reason	**Student**	**Dropout**	**Alt Completer**
Mental Exclusion			
Total	196 (60.0)	107 (32.7)	27 (7.3)
F2 Eligibility	107 (33.1)	38 (10.8)	3 (5.5)
F2 Ineligibility	73 (22.7)	63 (20.3)	4 (20.3)
Physical Exclusion			
Total	20 (84.4)	1 (5.1)	2 (10.4)
F2 Eligibility	11 (36.5)	1 (5.1)	1 (4.3)
F2 Ineligibility	8 (35.8)	--	1 (6.1)
Language Exclusion			
Total	130 (72.2)	38 (22.5)	8 (5.3)
F2 Eligibility	98 (53.7)	22 (12.8)	5 (3.7)
F2 Ineligibility	12 (6.1)	10 (5.3)	--

Table A-2
1988 Exclusion Reason by Race, Sex, Race and Sex, and 1992 Enrollment
(cont'd)

	Grade Progression Unwgt N (Wgt %)	
Eligibility/Exclusion Reason	**In Sequence**	**Out of Sequence**
Mental Exclusion		
Total	183 (54.9)	147 (45.1)
F2 Eligibility	97 (28.9)	69 (20.6)
F2 Ineligibility	72 (22.6)	68 (21.5)
Physical Exclusion		
Total	18 (80.2)	5 (19.8)
F2 Eligibility	9 (32.3)	4 (13.6)
F2 Ineligibility	8 (35.8)	1 (6.1)
Language Exclusion		
Total	120 (67.4)	56 (32.6)
F2 Eligibility	90 (49.8)	35 (20.5)
F2 Ineligibility	12 (6.1)	10 (5.3)

Table A-2
1988 Exclusion Reason by Race, Sex, Race and Sex, and 1992 Enrollment
(cont'd)

	Exclusion Reason Unwgt. N (Wgt.%)		
Race and Gender	**Physical**	**Mental**	**Language**
Asian-Female	--	1 (0.5)	29 (7.0)
Asian-Male	1 (0.1)	3 (0.4)	46 (6.1)
Hispanic-Female	--	20 (5.9)	46 (16.2)
Hispanic-Male	--	27 (5.2)	41 (9.1)
Black-Female	3 (2.7)	20 (11.3)	1 (0.5)
Black-Male	1 (0.9)	43 (13.6)	--
White-Female	6 (4.7)	53 (31.8)	2 (1.2)
White-Male	10 (3.4)	122 (43.6)	2 (0.5)
Am. Indian-Female	--	1 (0.7)	--
Am. Indian-Male	--	5 (1.8)	1 (0.4)
Unknown-Female	1 (0.8)	10 (5.1)	6 (2.3)
Unknown-Male	1 (0.3)	25 (7.2)	2 (0.3)

Table A-3
1992 Sample Ns of 1988 Excluded Students

Reason for 1988 Exclusion	Eligible	Ineligible	Out of Scope	Unlocat -able	1988 Sampled in Error*
Language Barrier	125	22	25	29	24
Physical Barrier	13	9	0	1	1
Mental Impairment	166	140	5	24	17
Unknown Reason	30	15	2	10	16
TOTAL	**334**	**186**	**32**	**64**	**58**

SOURCE: (for tables A-1, A-2, and A-3) -- NELS:88 Second Follow-Up, National Center for Education Statistics, U.S. Department of Education.

* Sampling error = sampled in error in the 1987-88 school year; either not a member of the eighth grade cohort, or no longer enrolled at the school at the time of sampling.

Appendix B:

Screener for Eligibility Reassessment

Appendix B:

NELS:88 First Follow-Up, Eligibility Criteria for the Base Year Ineligibles Study

The following instructions for the interpretation of eligibility were given to data collectors for the followback study of excluded 1988 eighth graders:

Eligibility criteria:

It is the intention of NELS:88 to include all sample members who are capable of meaningful participation in the regular survey under normal survey conditions. Therefore, UNLESS there are severe mental or physical handicaps or language barriers and the sample members are not capable of completing the survey instruments under normal circumstances of survey administration, the student should be considered ELIGIBLE for NELS:88; in cases where there is doubt of eligibility status, the sample member should be considered eligible.

It is extremely important that an individual assessment be made of each prospective sample member's ability to participate. Sample members are NOT to be considered ineligible categorically (for example, by virtue of being a special education student, or LEP student), since some special education students and many limited English-proficient students will be capable of completing the NELS:88 questionnaire or questionnaire and tests. In fact, as long as the sample member is capable of completing the NELS:88 Questionnaire, even if s/he is incapable of completing the NELS:88 Cognitive Test Battery, the sample member should be considered eligible for NELS:88.

Eligibility status will be determined by the school that the sample member attended during the 1989-90 school year (or last attended, in the case of dropouts). A knowledgeable official from this school must determine if a sample member is capable of participating in the regular NELS:88 survey (i.e., *a bilingual education/ESL/language arts teacher* for students with a language barrier, or *member of an Individualized Education Plan team* for physically or mentally handicapped students, or, exceptionally, other knowledgeable school staff who are personally familiar with the sample member's performance). In order to assist schools in making eligibility determinations, we urge that the following guidelines be followed:

1. Limited English Proficiency (LEP) and No English Proficiency (NEP) students.

A non-native speaker of English who has been in an English-language course of study less than two years may be considered ineligible if the school determines that the sample member would not be capable of completing a NELS:88 questionnaire; this study assumes that, normally, sample members who have been enrolled in an English-language course of study for at least two years should be considered eligible for NELS:88.

However, a sample member who is a native speaker of Spanish who is literate in Spanish should be included, even if that student is an NEP. Any native speaker of Spanish should be considered eligible for NELS:88 if s/he is capable of completing either the English language OR the Spanish-language version of the NELS:88 Questionnaire.

2. Special Education students.

To be considered a Special Education student, the student should normally have an Individualized Education Plan (IEP). A student with an IEP (or dropout with an IEP before dropping out)--who is not mainstreamed in English/language arts; AND who is judged by the school to NOT be capable of completing the NELS:88 Questionnaire should be considered ineligible for NELS.

To guide schools in making this determination, we recommend that a sample member be judged incapable of completing the NELS:88 Questionnaire if that sample member READ AT A GRADE LEVEL FOUR OR MORE LEVELS BELOW THE TENTH GRADE NORM during the 1989-90 school year in English (or Spanish). This may be determined by looking at school records such as standardized test scores, or, particularly when these are not available, by the judgment of the language arts teacher, IEP team members, or other school staff familiar with the sample member's performance.

3. Students with behavioral disorders or severe physical impairments.

A sample member who is not normally subject to testing or surveying by the school, owing to behavior disorders, severe cognitive deficits, or severe physical impairments which do not permit the testing/surveying of the sample member under any but extraordinary conditions, should be considered ineligible for NELS:88.

"Extraordinary conditions" should be defined in relation to departure from normal testing and survey protocols. That is, if the sample member would require oral rather than self-administration of the survey instruments, large print or braille versions of the survey forms, translations other than in Spanish, or other extraordinary special assistance or aids, that sample member should be deemed ineligible for participation in NELS:88.

NELS:88 2FU: FOLLOWBACK STUDY OF EXCLUDED STUDENTS **NORC 4521**
STUDENT ENROLLMENT STATUS SCREENER 3/92
HOME CONTACT

Use this screener to:

1. verify the enrollment status and confirm the dropout/dropin dates of sample members reported to be **Phase One-Two dropouts**. **START AT QUESTION 1.**
2. verify the enrollment status and confirm the dropout/dropin dates of sample members reported to be **Phase Three dropouts**. **START AT QUESTION 7. (IF A PHASE ONE-TWO AND PHASE THREE DROPOUT, START AT QUESTION 1.)**
3. determine the current enrollment status of and the school most recently attended by BYI's who were not enrolled in school as of the First Follow-Up. **START AT QUESTION 7.**

DATE: / /92

FI NAME: _____ FI ID: |___|___|___|___|___|___|

STUDENT ID: |___|___|___|___|___|___|___|___|

STUDENT NAME: _____

CONTACT NAME: _____

IF NOT SAMPLE MEMBER, RELATION TO SAMPLE MEMBER: _____

ASK TO SPEAK TO THE STUDENT (OR PARENT/GUARDIAN--MODIFY QUESTIONS AS NECESSARY)

A. Hello, this is _____ calling from the National Opinion Research Center at the University of Chicago. I'm calling regarding the Second Follow-Up to the National Education Longitudinal Study of 1988. We are currently following up on some of our sample members, and I'd like to ask you a few questions about your enrollment in school.

1. **Were you out of school for a month or more for a reason other than illness or vacation between March 1, 1991 and December 15, 1991?**

   Yes . . . . . 1 **SKIP TO QUESTION 4** (SAMPLE MEMBER IS A PHASE ONE-TWO DROPOUT)
   No . . . . . 2 **GO TO QUESTION 2**

2. **Your school records from (SCHOOL) indicate that you first left that school in (DROPOUT DATE REPORTED BY SCHOOL). Did you transfer to another school, and if so, when?**

   Yes . . . . . 1 --> **TRANSFER DATE**: |___|___|-19|___|___| --> **GO TO QUESTION 3**
   No . . . . . 2 **RECORD BELOW WHAT THE STUDENT WAS DOING, THEN SKIP TO Q7**

_____

_____

_____

3. What is the name and location of the school to which you transferred?

NAME OF SCHOOL: _____ PIN: |__|__|__|__|__|__|__|

CITY: _____ STATE: _____

SAMPLE MEMBER IS A PHASE ONE-TWO STUDENT. SKIP TO QUESTION 7.

4. During which month and year did you first leave school for a month or more for reason other than illness or vacation?

|__|__| - 19 |__|__| **(EARLIEST DROPOUT DATE)**
MONTH YEAR

5. Did you return to school, and if so, during which month and year?

Yes . . . . . 1 **DROP-IN DATE:** |__|__| - 19 |__|__| **--> GO TO QUESTION 6A**
MONTH YEAR

No . . . . . 2 **SKIP TO QUESTION 8** (SAMPLE MEMBER IS A PHASE THREE DROPOUT)

6A. Did you leave school again for a month or more at some later point, and if so, when?

Yes . . . . . 1 **DROPOUT DATE:** |__|__| - 19 |__|__| **--> GO TO QUESTION 6B**
MONTH YEAR

No . . . . . 2 **SKIP TO QUESTION 8** (SAMPLE MEMBER IS A PHASE THREE STUDENT)

6B. Did you return to school, and if so, during which month and year?

Yes . . . . . 1 **DROP-IN DATE:** |__|__| - 19 |__|__| **--> COLLECT REMAINING DROPOUT/IN DATES, THEN SKIP TO QUESTION 8.**
MONTH YEAR

No . . . . . 2 **SKIP TO QUESTION 8** (SAMPLE MEMBER IS A PHASE THREE DROPOUT)

7. Are you currently attending school?

Yes . . . . . 1 **GO TO QUESTION 8**

No . . . . . 2 **GO BACK TO QUESTION 4** (SAMPLE MEMBER IS A PHASE THREE DROPOUT)

8. What is the name and location of the school you are currently attending/most recently attended?

NAME OF SCHOOL: _____ PIN: |__|__|__|__|__|__|__|

CITY: _____ STATE: _____

NOW GO TO PARAGRAPH B

B. Thank you very much for your assistance. Another representative from the National Opinion Research Center may be contacting you shortly about NELS:88.

NELS:88 2FU: FOLLOWBACK STUDY OF EXCLUDED STUDENTS **NORC 4521**
STUDENT ENROLLMENT STATUS SCREENER 3/92
SCHOOL CONTACT

DATE: / /92

FI NAME: _____ FI ID: |___|___|___|___|___|___|

STUDENT ID: |___|___|___|___|___|___|___|

STUDENT NAME: _____

SCHOOL NAME: _____

CONTACT NAME: _____

ASK TO SPEAK TO A GUIDANCE COUNSELOR

Hello, this is _____ calling from the National Opinion Research Center at the University of Chicago. I'm calling regarding the Second Follow-Up to the National Education Longitudinal Study of 1988. We are currently following up on students who were excluded from the student sample in 1988 or 1990 because of a language barrier or physical or mental disability. We are attempting to locate the students to determine whether they are eligible for the survey. We think that one/some of these students may be enrolled in your school, and I would like to ask you a few questions about him/her/them.

1. Is (STUDENT) currently enrolled in your school?

Yes	1	**SKIP TO QUESTION 4**
No, he/she dropped out (has had 20 or more consecutive unexcused absences)	2	**SKIP TO QUESTION 3A**
No, he/she transferred to another school	3	**GO TO QUESTION 2**
No -- OTHER (home-study, early graduate, institutionalized, etc.) SPECIFY BELOW	4	**SKIP TO QUESTION 3A**

_____

2. What is the name and location of the school to which the student transferred?

NAME OF SCHOOL: _____ CITY _____ STATE _____

GO TO QUESTION 3A

3A. In what grade was (STUDENT) when he/she left your school?

GRADE: |___|___|

Not assigned a grade-level 1

3B. When did he/she last attend your school?

¦___¦___¦ - 19 ¦___¦___¦ --> **IF PRIOR TO MARCH 1, 1991, STOP. SKIP TO BOX AT BOTTOM OF NEXT PAGE. (IF DROPOUT, THIS DATE IS MOST RECENT DROPOUT DATE.) IF AFTER MARCH 1, CONTINUE WITH QUESTION 4.**
MONTH YEAR

4. Did (STUDENT) have 20 or more consecutive unexcused absences between March 1, 1991 and December 15, 1991?

Yes . . . . . 1 **GO TO QUESTION 5A**

No . . . . . 2 **SKIP TO BOX AT BOTTOM OF PAGE**

Not enrolled during that time period . . . . . 3 **SKIP TO BOX AT BOTTOM OF PAGE**

5A. During which month and year did he/she first stop attending school?

¦___¦___¦ - 19 ¦___¦___¦ **(DROPOUT DATE)**
MONTH YEAR

5B. In what grade was (STUDENT) at that time?

GRADE: ¦___¦___¦

Not assigned a grade-level 1

6. Did he/she return to school, and if so, during which month and year?

Yes . . . . . 1 **DROP-IN DATE**: ¦___¦___¦ - 19 ¦___¦___¦ **--> GO TO QUESTION 7A**
MONTH YEAR

No . . . . . 2 **SKIP TO BOX AT BOTTOM OF PAGE**

7A. Did he/she stop attending school again at some later point, and if so when?

Yes . . . . . 1 **DROPOUT DATE**: ¦___¦___¦ - 19 ¦___¦___¦ **--> GO TO QUESTION 7B**
MONTH YEAR

No . . . . . 2 **SKIP TO BOX BELOW**

7B. In what grade was (STUDENT) at that time?

GRADE: ¦___¦___¦

Not assigned a grade-level 1

COLLECTING REMAINING DROPOUT/IN DATES, IF NECESSARY. THEN GO TO BOX BELOW

SCREEN FOR ENROLLMENT FOR EACH EXCLUDED STUDENT REPORTED TO ATTEND THE SCHOOL. THEN ADMINISTER THE ELIGIBILITY SCREENER FOR EACH STUDENT WHO IS REPORTED TO BE A DROPOUT, AN ENROLLED STUDENT, OR AN "OTHER" IN QUESTION 1.

DO NOT ADMINISTER THE ELIGIBILITY SCREENER FOR STUDENTS WHO HAVE TRANSFERRED TO ANOTHER SCHOOL. YOU WILL INSTEAD CONTACT THE NEW SCHOOL AND ADMINISTER THE ENROLLMENT AND ELIGIBILITY SCREENERS FOR THAT STUDENT.

NELS:88 2FU: FOLLOWBACK STUDY OF EXCLUDED STUDENTS **NORC 4521**
ELIGIBILITY SCREENER 3/92

DATE: / /92

FI NAME: _____ FI ID: |__|__|__|__|__|__|__|

STUDENT ID: |__|__|__|__|__|__|__|__|

STUDENT NAME: _____

SCHOOL NAME: _____

CONTACT NAME: _____

Now I would like to ask a few questions that will help us determine whether (STUDENT) is eligible for the Second Follow-Up of NELS:88. **NOTE: IF THE STUDENT IS NOT CURRENTLY ENROLLED IN THE SCHOOL, YOU WILL NEED TO MAKE THE APPROPRIATE CHANGES IN VERB TENSE IN THE SCREENER QUESTIONS, AND MAY NEED TO SUBSTITUTE "MOST RECENT" FOR "1991-92' IN QUESTIONS 1 AND 5.**

1. Was (STUDENT'S) 1991-92 (most recent) reading score in ENGLISH below the eighth grade level?

Yes	1	**GO TO QUESTION 2**
No	2	**GO TO QUESTION 2**
Unable to assign a grade level to score	3	**GO TO QUESTION 2**
No reading score available from 91-92	4	**SKIP TO QUESTION 3**

ALSO RECORD ANSWER, VERBATIM _____

_____

2. What was that reading score? (WRITE IN BELOW)

_____

IF BELOW EIGHTH GRADE, SKIP TO Q4. IF EQUAL TO OR HIGHER THAN EIGHTH GRADE, SKIP TO Q9. IF UNABLE TO ASSIGN A GRADE LEVEL, GO TO Q3.

3. Do you feel that (STUDENT) is capable of completing a questionnaire designed for students who read English at the eighth grade level?

Yes	1
No	2

ALSO RECORD ANSWER, VERBATIM _____

_____

_____

_____

4. Is (STUDENT) a native speaker of Spanish?

Yes	1	**GO TO QUESTION 5**
No	2	**SKIP TO QUESTION 8**

5. Was his/her 1991-92 (most recent) reading score in Spanish below the eighth grade level?

Yes	1	**GO TO QUESTION 6**
No	2	**GO TO QUESTION 6**
Unable to assign a grade level to score	3	**GO TO QUESTION 6**
No reading score available from 91-92	4	**SKIP TO QUESTION 7**

ALSO RECORD ANSWER, VERBATIM _____

_____

6. What was that score? (WRITE IN BELOW)

_____

IF BELOW EIGHTH GRADE, SKIP TO Q9. IF EQUAL TO OR HIGHER THAN EIGHTH GRADE, SKIP TO Q9.
IF UNABLE TO ASSIGN A GRADE LEVEL TO THE SCORE, GO TO Q7.

7. Do you feel that (STUDENT) is capable of completing a questionnaire designed for students who read Spanish at the eighth grade level?

Yes	1
No	2

ALSO RECORD ANSWER, VERBATIM _____

_____

NOW SKIP TO QUESTION 9

8. Is (STUDENT) a native speaker of any language other than English or Spanish?

Yes	1	--->	**What language?** _____
No	2		

9. Does (STUDENT) suffer from any behavior disorder, severe cognitive deficit, or severe physical impairment that would make it necessary for one to use extraordinary measures to administer the questionnaire, such as oral administration versus self-administration, a large print or Braille version of the questionnaire, or other extraordinary special assistance?

Yes	1	**GO TO QUESTION 10**
No	2	**SKIP TO QUESTION 12**

10. Please specify the behavior disorder, severe cognitive deficit, or physical impairment:

RECORD ANSWER, VERBATIM _____

_____

_____

11. Please specify the extraordinary measures that would be needed to administer the questionnaire:

RECORD ANSWER, VERBATIM _____

_____

_____

_____

12. In what grade is (STUDENT)?

GRADE: |___|___|

Not assigned a grade-level 1

13. What is (STUDENT'S) sex?

Male 1

Female 2

14. Of the following choices, which best describes (STUDENT)? (CIRCLE ONE)

Asian or Pacific Islander 1
Hispanic, regardless of race 2
Black, not of Hispanic origin 3
White, not of Hispanic origin 4
American Indian or Alaskan Native . . . 5

15. What are his/her home phone number, address, and parents' names, according to your most recent records?

ADDRESS _____

CITY _____ STATE _____ ZIP CODE |__|__|__|__|__|

MOTHER/FEMALE GUARDIAN'S NAME _____

FATHER/MALE GUARDIAN'S NAME _____

PHONE NUMBER (_____)_____

16. Finally, I would like to make sure that I have the correct address for your school.

ADDRESS _____

CITY _____ STATE _____ ZIP CODE |__|__|__|__|__|

PHONE NUMBER (_____)_____

ADMINISTER ELIGIBILITY SCREENER FOR NEXT STUDENT, OR, IF NO MORE STUDENTS, THANK CONTACT.

Appendix C:

HS&B and NAEP Eligibility Criteria

Appendix C:
HS&B Baseline (1980), NAEP (1990, 1992), Eligibility Criteria

C-I: HS&B. NELS:88 base year eligibility rules were modelled after those of High School and Beyond (HS&B), the prior NCES school-based longitudinal survey. Comparability across the studies was a major aim, since NELS:88 was designed to monitor trends by providing a basis for cross-cohort comparisons with the two (senior, sophomore) 1980 HS&B cohorts.

In High School and Beyond, the student sample frame was derived from the eligible school sample. (As in NELS:88, certain types of special schools were excluded.) Within the eligible schools, a sophomore was defined as a student who expected to complete his/her tenth grade course work between April 1, 1980 and August 31, 1980. This was to include those students who might be held back or who might repeat tenth grade, but to exclude students dropping out before administration of the HS&B questionnaire in the spring of 1980.

In HS&B, a senior was defined as a twelfth grade student who expected to complete his/her high school course work between April 1, 1980 and August 31, 1980. This group included students who might repeat the grade, as well as "early completers" if they were to complete their course work during this time period. The twelfth grade cohort, however, was not to include early or late graduates.

NORC asked each selected school to provide a list of its tenth and twelfth grade students, as defined above. All students defined as being eligible for sample selection were included in the sampling frame; conversely, all students defined as ineligible for sample selection were removed from the frame.

Additionally, however, a student was considered ineligible when that student:

- Was a foreign exchange student
- Transferred out of the selected school. (A transfer student was defined as a student who had left the school and whose records were requested for a new school).[1]
- Died.
- Would be unavailable until after August 31, 1980.
- Was listed on the roster in error.

[1] Transfers-out were not directly replaced. (However, as in NELS:88, all transfers-in were given a chance of selection into the sample). All other categories above in HS&B base year led to replacement by other students from the roster. (HS&B substituted students for the following cases: dropout, listed in error, language barrier, too ill [mentally, physically], in jail, unavailable entire field period, expelled, and deceased.) No substitution of students was done in the NELS:88 base year.

- Had become a drop-out or lost student since he was selected. Such a student would have to have been out of school for 20 or more consecutive days and was not expected to return.[2]

- Was physically or mentally unable to participate in the survey.[3]

However, there were a number of weaknesses in HS&B's approach to student exclusion. Among these is the fact that numbers and characteristics of excluded IEP and LEP students were not clearly documented.

While an overall exclusion rate is not reported in the HS&B documentation, there is a good deal of evidence that students with (for example) learning disabilities are largely absent from the HS&B sample. Hoachlander (1992) notes that "according to Harnisch, Lichtenstein, and Langford [Harnisch, D.L., Lichtenstein, S., Langford, J.B., 1986, *Digest on Youth in Transition,* Champaign, Illinois] 94 percent of the students who can be positively identified as handicapped in HS&B were physically handicapped; the national rate of physical disabilities among school-age children with special needs is 4 percent. Only 6 percent of the students identified as handicapped in the HS&B sample were learning disabled, and none ere emotionally disabled or retarded. The vast majority of all handicapped students is generally comprised of these three disability groups, so the sample of handicapped students in HS&B . . . is in no way representative of the national population of handicapped students."

For those students with disabilities who are included in the HS&B sample, a methodological monograph by Owings and Stocking (1985) examines the stability and implications of self-report of disability status.

HS&B oversampled Hispanic students, but did not oversample Asians or others who might be more likely to have large LEP populations. In order to minimize the number of LEP students to be excluded, HS&B provided for a Spanish-language version of the 1980 sophomore and senior questionnaires. However, the low number of individuals who elected to complete Spanish language instruments does not encourage belief that all LEPs were included.

---

[2] Dropouts meet the 20 consecutive days criterion, are at least 16 years of age, and are not expected to return to school. Lost students are dropouts in all respects except that they are not 16 years of age.

[3] While this category was used to cover linguistic exclusion also, a Spanish language version of the questionnaire was provided so that students whose primary language competence was in Spanish would not be excluded. However, only 43 sophomores and 13 1980 HS&B seniors elected to complete the instrument in Spanish (*High School and Beyond Information for Users, Base Year 1980 Data*, Chicago, NORC, December 1980, p.8-3 Item 9).

C-II: National Assessment of Educational Progress (NAEP).

Although NAEP had long excluded students whose schools deemed them incapable of meaningful participation, for the 1990 assessment, NAEP moved toward greater specificity in supplying objective criteria for schools to follow in making eligibility determinations. Schools were provided with the following criteria for the 1990, 1992 and 1994 assessments:

National Assessment of Educational Progress
Criteria for Excluding Student from the Assessment

The intent is to assess all selected students. Therefore, all sampled students who are capable of participating in the assessment should be assessed.

Some students may be incapable of participating meaningfully in the assessment because of limited English proficiency or a physical or mental handicap. The Local Administrator, with the advice of other staff members, may exclude from the assessment only those students who are incapable of taking the assessment because:

- The student is a native speaker of a language other than English and has been enrolled in an English-speaking public school (not including a bilingual education program) for less than two consecutive years;

OR

- The student is a special education student with a Individualized Education Plan (IEP) who is mainstreamed less than 50 percent of the time in academic subjects and the IEP team has determined that the student is unable to be assessed.

Students with limited English proficiency and students with IEP's should be assessed, if in the judgment of school staff, they are capable of taking the assessment. When there is doubt, include the student.

Undoubtedly these criteria, representing as they do a move from vague and quite general to more specific decision rules, represent an improvement on past practice. Nevertheless, these guidelines have been criticized as still leaving much to be desired. For example, Ysseldyke, Thurlow, McGrew and Vanderwood (1994), in their summary of the recommendations and conclusions of a working conference to develop guidelines on inclusion and accommodations, note that:

> The current NAEP guideline for making inclusion/exclusion decisions is problematic in at least two ways. One way is its use of a percentage of time in the mainstream setting. The second way is in its reliance on the "IEP team" (or some designated person) to make decisions about "meaningful participation" in assessments.

(a) Percentage of time in the mainstream is not a good indicator of a student's instructional program, level of skill development, or ability. There are too many other factors that enter into mainstreaming decisions.

Rather than using a percentage of time measure, a better indicator would be correspondence between the content the test is intended to measure and the type of curriculum for the students. Students who are working toward outcomes other than those measured by the assessment (e.g., functional skills) should participate in an alternative form of assessment. The type of curriculum rather than the setting should be the factor that determines the nature of the assessment.

(b) The "IEP team" allows too much slippage in the team decision-making process. Frequently, the IEP terminology is interpreted to mean that any student on an IEP should be excluded from testing. Sometimes decisions are made solely on the basis of the student's category of disability. It is not a good idea to encourage IEP teams to decide whether students should take tests.

Rather than referring to the IEP, it would be better to identify skills needed to take the assessment. School building administrators could be provided with a checklist of factors to consider in making inclusion/exclusion decisions.

The impact of exclusion rules can be seen in results for the 1992 mathematics assessment reported below.

Table C-1:
Weighted Percentages of Fourth and Eighth Grade Students Excluded (IEP and LEP) from NAEP 1992 Public School Sample

	Nation	Northeast	Southeast	Midwest	West
Grade 4					
Percent IEP, LEP	12	12	11	7	18
Percent Excluded	8	8	7	5	12
Percent IEP	9	9	9	6	10
Percent IEP Excluded	6	5	6	4	6
Percent LEP	4	3	1	1	9
Percent LEP Excluded	3	3	1	1	7
Grade 8					
Percent IEP, LEP	12	12	11	9	15
Percent Excluded	7	8	6	6	9
Percent IEP	9	10	10	8	8
Percent IEP Excluded	5	6	5	5	5
Percent LEP	3	3	1	1	8
Percent LEP Excluded	2	2	1	1	4

Source: NCES, National Assessment of Educational Progress; Adaptation of Table B.4, Mullis et al. 1993, ***NAEP 1992 Mathematics Report Card for the Nation and the States.***

Note: Students reported as both IEP and LEP counted once in overall rate and separately for LEP and IEP.

1995 Field Test Revised Eligibility Criteria

NAEP experimented with new criteria in the 1995 field test. For the 1995 field test of NAEP, criteria were broadened specifically to encourage greater inclusion of students with disabilities or limited English proficiency by providing some specific modifications and accommodations which it was a purpose of the field test to evaluate. The 1995 criteria were as follows:

A student with limited English proficiency will be included unless:

- he or she is a native speaker of a language other than English, and
- the student has received academic instruction primarily in English for less than three school years, and
- school officials judge the student to be incapable of taking the assessment.
- in a sample of schools the field test in mathematics included use of test booklets in Spanish, or test booklets with facing pages of Spanish and English, for use with Spanish-speaking students whose proficiency in English was limited.

A student with disabilities will be included if:

- the student's IEP team or equivalent group has determined that the student cannot participate in tests such as NAEP, or
- if the student's cognitive functioning is so severely impaired that he or she cannot participate even with accommodations.

In the 1995 field test, accommodations specified by the student's IEP were provided to students sampled for mathematics sessions. These accommodations included extended time, oral administration, Braille test booklets, large-print test booklets, talking calculators, one-on-one administration possibly in a separate location, and other accommodations as specified by the student's IEP.

As in the past, where doubt existed as to the student's ability to participate in the regular NAEP, for both LEP and IEP students, the student was to be included.

Appendix D:

Technical Notes

Appendix D
Technical Notes

Analysis Methodology

The eighth grade cohort expanded sample file was used for this analysis. It comprises 20,070 cases, including 584 base year ineligibles. The expanded eighth grade cohort comprises a subsample of the base year eligible and ineligible eighth graders. Subsampling of the eligible eighth grade cohort took place as part of the regular first follow-up survey. A subsample of ineligible 1988 eighth graders was also selected at that time. Excluded from the analysis file were cases originally classified as base year ineligibles but later found to be sampling errors, and cases that were out of scope for the second follow-up (deceased, out of country).

The base for the expanded cohort is not absolutely stable because sample members can become permanently out of scope through death, or be temporarily out of scope through expatriation. For example, a 1988 eighth grader may have been outside the United States in 1990 at the time of the first follow-up. That sample member would be classified as out of scope for the 1990 round. If that sample member remained outside the United States for the 1992 round, the sample member would again be out of scope. If, however, the sample member returned to the United States, that individual was considered in scope for the 1992 follow-up.

The N for the (subsampled) expanded cohort is 20,263 for the base year. Owing to out-of-scopes, the N is 20,133 for the first follow-up, and 20,070 for the second follow-up. Figure D-1 illustrates the status of the base year expanded cohort across rounds.[1] With the addition to the expanded sample file of freshened sophomore cohort spring 1990 eligibles and ineligibles and freshened senior cohort spring 1992 eligibles, there are 21,133 cases on the 1992 expanded sample file (not counting deceased sample members and out-of-scopes).

The expanded sample weight (F2EXPWT) was used with this analysis. (Statistical properties of this weight are documented in the forthcoming *NELS:88 Base Year Through Second Follow-Up Sampling Design, Weighting and Estimation Report*; Ingels, Scott & Frankel, NCES 1996). While most analyses employed the sample weight, for some purposes it was preferable to report unweighted proportions. The tables in this report clearly identify whether weighted population estimates or unweighted sample Ns or proportions are presented. Also employed in this analysis were the following variables, which appear on the expanded sample file:

BYISAMP. Value "1" identifies members of the base year ineligibles sample.

G8REGON. Values encompass the northeast, midwest, south, and west. Region is based on the location of the eighth grade school (1988).

[1] The expanded sample file, however, reflects a total of 21,133 eligible and ineligible sample members. This is the case because it contains records both for the 1990-retained 1988 eighth grade cohort analyzed in this report, *and* the (overlapping) 1990 sophomore cohort. In the 1990 first follow-up, additional individuals were added to the sample from freshening, some candidates for freshening were classified as ineligible, some eligible sample members became ineligible owing to accidents or illnesses that gravely impaired them, and still other members of the 1990-retained eighth grade cohort did not qualify for membership in the sophomore cohort because they were not tenth graders in 1990.

F2RACE. This variable is based on race/ethnicity information available as of the 1992 second follow-up. Sample members were categorized in terms of the following values: unknown, Asian, Hispanic, Black, White, American Indian.

F2SEX. This variable is based on the NELS:88 second follow-up and the values are "male" and "female".

BYINELR. This variable contains information about the reason for ineligibility in the NELS:88 base year. Values include: ineligibility reason unknown, physical disability, mental disability, language barrier (LEP).

ENRL2EXP. This variable describes enrollment status as of the second follow-up of NELS:88 (spring 1992). Values are: status unknown, student, alternative completer (dropped out but receiving alternative instruction, e.g. for GED), dropout.

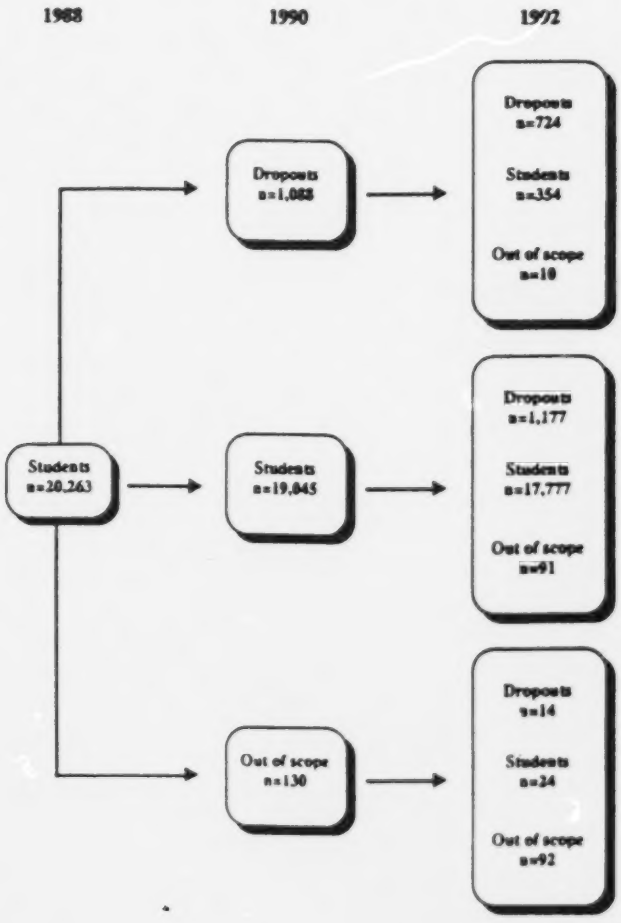
SEQSTAT. This variable indicates whether the sample member is or is not a spring 1992 high school senior.

F2INELST. This variable indicates the 1992 enrollment status of 1988-ineligible eighth graders. Statuses for base year ineligibles are as follows: second follow-up eligible; second follow-up ineligible; second follow-up unlocatable; second follow-up out of scope.

NELS:88 data users can identify eligible and ineligible members of the eighth-grade and sophomore cohorts through the variables G8COHEXP and G10COHEXP. Members of the base year ineligible sample can be identified through the variable BYIXREAS (this variable has four values, standing for "Reason for Ineligibility in the Base Year").

The expanded sample file is available on the 1995 privileged use ECB release of the NELS:88 data. Contents of the file and other information for users in provided in the *NELS:88 Base Year Through Second Follow-Up Final Methodology Report* (Ingels, Scott, Taylor & Moore; NCES, 1996).

Figure D-1:
Status of NELS:88 8th Grade Expanded (Eligible + Ineligible) Cohort Sample as of Base Year (1988), First Follow-Up (1990), and Second Follow-Up (1992)



Notes:

1) The tree includes all base-year eligible and ineligible eighth graders who were retained after first follow-up subsampling and who were students (regardless of whether they were in tenth grade) or dropouts in the first follow-up. The 1990 in-scope N = 20,133; the 1992 in-scope N = 20,070.

2) "Out of scope" includes deceased and out of country. Sample members who are out of the country are considered to be temporarily out of scope. If they have returned at the time of the next data collection, they are considered in scope for that data collection.

3) "Students" includes students enrolled in high school, early graduates, and dropouts who have earned a GED or who are receiving any sort of alternative academic instruction.

4) As part of the process for generation of the second follow-up expanded sample weight, enrollment status was imputed when 1990 or 1992 status was unknown (i.e., for questionnaire noncompleters). These imputed values are used in the tree. Hence there are no sample members with "unknown" enrollment status.

Study Methodology

First Follow-Up Survey of Base Year Ineligible Students. The Base Year Ineligibles (BYI) Study of the NELS:88 first follow-up was a followback of students who had been excluded because of linguistic, mental, or physical obstacles to participation when the baseline sample of eighth graders was drawn in the 1987-88 school year. The BYI study had several purposes, the primary foci of which were to correct for potential sample undercoverage; to accommodate the group of 1988-ineligible sample members who were 1990-eligible sophomores, and hence must be added to the 1990 survey to ensure its cross-sectional representativeness; and to provide a basis for a corrected cohort dropout estimate taking account of both 1988-eligible and 1988-ineligible eighth graders two years later.

Two kinds of information were sought from the sample of excluded students. First, it was to be determined if their eligibility status had changed. If so, these students were to be reclassified, and added to the longitudinal sample. They would then be administered, as appropriate, a student or dropout questionnaire. Second, for those who remained ineligible, their school enrollment status was to be ascertained, and basic information about their sociodemographic characteristics recorded. For eligibility and completion rate data, and further details, see the *NELS:88 First Follow-Up Final Technical Report* (Ingels, Scott, Rock, Pollack, & Rasinski, NCES 1994).

Followback Study of Excluded Students (FSES). Of the 618 base year ineligible sample members (BYIs), 580 were located and 312 were reclassified as eligible during the first follow-up. In the second follow-up, the remaining ineligible students were pursued.

The Followback Study of Excluded Students of the NELS:88 second follow-up attempted to reassess the eligibility status and ascertain the enrollment status of students who: 1) had been excluded because of linguistic, mental, or physical obstacles to participation when the baseline sample of eighth graders was drawn in the 1987-88 school year, and were subsampled into the Base Year Ineligible Study in the first follow-up; 2) were eligible in the base year but became ineligible in the first follow-up; or, 3) were identified as ineligible when selected through the freshening process in the first follow-up. If the students had since become eligible for NELS:88, the followback study attempted to survey them.

Data collection for the followback study of base year excluded students took place between April and October, 1992. Interviewers attempted to identify excluded students who were eligible to be added to the longitudinal sample in the second follow-up. They obtained the following information about the excluded student from the student's current school, school last attended, or the student's home:

- **Sex (if unknown):** male or female;
- **Race/ethnicity (if unknown):** white, black, Hispanic, Asian/PI, American Indian, other;
- **School enrollment status:** student, dropout, or dropout in alternative program; and,
- **Eligibility:** English/Spanish language proficiency, lack of mental or physical disability (i.e., ability to complete a questionnaire), reading ability level of at least eighth grade.

After collecting the above information about the students, interviewers attempted to identify whether or not the student was capable of meaningful participation in the survey under normal conditions. To make this assessment, interviewers were instructed to obtain reports from persons with first-hand

knowledge of the students, such as a special education teacher, a bilingual teacher, a language arts teacher, or a guidance counselor. Interviewers often spoke with several staff members to identify the staff member who was most qualified to assess whether or not the student could participate. Unless there were severe mental or physical disabilities or lack of facility with written English or Spanish and the member was unable to complete the survey instruments under normal circumstances, the student was considered eligible to participate in the study.

The results of data collection for FSES are summarized in Table 4.3.6-1 of the *NELS:88 Second Follow-Up Student Component Data File User's Manual* (Ingels, Dowd, Baldridge, Stipe, Bartot, and Frankel, 1994, NCES 94-374). Eligibility information was gathered for 94.7 percent of the excluded sample members. For excluded students who were identified as eligible, student or dropout questionnaires were administered either in-person or over the telephone. Cognitive tests were administered in addition to the questionnaires in some cases, when an in-person questionnaire administration took place. Of the 239 base year ineligibles who were eligible and surveyed in the NELS:88 second follow-up, 88 completed a cognitive test in addition to a questionnaire and 151 completed a questionnaire only. The low proportion of test completions primarily reflects the high costs of these cases, resource limitations, and frequent use of telephone data collection methodology. For students who remained ineligible, school enrollment status and information on other key characteristics were obtained.

Using the Ineligibles Sample Data. Beginning with the second follow-up release, 1988 data for ineligible sample members who were reclassified as eligible in 1990 or 1992 and completed a questionnaire (or, for 1992, a questionnaire and the NELS:88 achievement battery) have been included on the files. Thus 1990 (and 1992) cross-sectional estimates reflect participating reclassified students who had been excluded in 1988. These cases are automatically invoked in cross-sectional analysis through use of the sophomore and senior cohort flags, providing, of course, that the 1988-excluded student was an eligible, participating sophomore in 1990 or senior in 1992.

Limited longitudinal analyses with questionnaire data are possible, given that excluded students who were eligible, participating 1990 sophomores or 1992 seniors were potentially eligible for panel weights in the second and third follow-up. Panel weights based on eighth grade cohort status are not provided. If a 1988 ineligible eighth grader was reclassified and participated in 1990 and 1992 and was a sophomore in 1990, that individual will automatically fall into the sample for 1990 to 1992 sophomore cohort panel analyses. Reclassified excluded students are also represented in the third follow-up (1994) sample. Hence these individuals contribute to sophomore cohort estimates for 1990 to 1994 as well as to longitudinal (1992-1994) estimates for the senior cohort. However, such an individual, having no base year questionnaire data, would not have been assigned a 1988 to 1990 or 1988-1990-1992 eighth grade cohort panel weight. The only panel analyses that can be conducted with ineligible eighth graders that take the 1988 base year as their starting point are analyses using the expanded cohort variables and the expanded cohort weight.

The eighth grade cohort expanded sample with the corresponding expanded sample weight (F2EXPWT) is designed to be used only with the variables on the expanded sample file. While questionnaire data are available for eligible sample members from the student or dropout questionnaires, such data are systematically missing for all ineligible sample members. Because of the many non-random differences between eligible and ineligible sample members, invocation of student and dropout questionnaire data will produce biased estimates for the expanded sample. The available expanded cohort variables are limited--cohort membership, 1990 and 1992 enrollment status, whether in modal grade

sequence in 1990 or 1992, reason for exclusion in 1988, sex, race/ethnicity, school type (1988, 1990, 1992), school's metropolitan status (urbanicity as of 1988, 1990, and 1992), school's census region (1988, 1990, 1992). This limited information is designed to support calculation of basic cohort dropout rates (see, for example, Kaufman, McMillen & Whitener, 1991, *Dropout Rates in the United States: 1990*, NCES 91-053, pp. 16-18) and methodological analyses such as offered in this report.

Some information for excluded sample members also appears in the NELS:88 transcript data. Base year excluded eighth graders were eligible for inclusion in the transcript data collection if they had been reclassified as eligible by 1992 and were attending an institution within the subset of NELS:88 schools from which academic transcripts were collected. In addition, transcripts were collected for excluded eighth graders who remained ineligible in 1992 but who were 1992 seniors. This was so that the cross-sectional 1992 senior cohort transcript sample could more nearly approximate and be comparable to the NAEP 1990 and 1994 high school transcript samples. Additional enrollment status variables, based on records sources, are also supplied in the transcript data, as well as a status variable comparing school records and survey-based enrollment status reports.[2]

[2] For further details, see Ingels, Dowd, Taylor, Bartot, Frankel & Pulliam, 1995, *NELS:88 Second Follow-Up: Transcript Component Data File User's Manual*, Washington, D.C.: NCES 95-377.

Appendix E:

Glossary

Appendix E
Glossary

Alternative assessment: see "Performance assessment".

Alternative completer: The NELS:88 second follow-up distinguished three levels of enrollment status: students enrolled in a regular high school program, dropouts who had enrolled in (or had completed) some alternative (non-diploma) high school equivalency accrediting program (for example, preparation classes for the GED test), and dropouts receiving no alternative instruction. The term "alternative completer" was used for dropouts receiving any sort of instruction to prepare them for equivalency certification, and for dropouts who had already received the GED or other equivalency certification.

Base year ineligible (BYI) study: A NELS:88 First Follow-Up study which sought to locate and survey eligible respondents who were part of the Base Year sample, yet were ineligible to participate in the Base Year due to mental or physical incapacity, language barrier, or other factors. (See entry for "Followback study of excluded students.")

Bias (due to nonresponse): Difference that occurs when respondents differ as a group from nonrespondents on a characteristic being studied.

Bias (due to undercoverage): This bias arises because some portion of the potential sampling frame is missed or excluded. For example, if the school list from which a school sample is drawn is incomplete or inaccurate, school undercoverage may occur. In NELS:88 the most important potential source of undercoverage bias was exclusion of 5 percent of the potential sample of eighth graders in the base year. (See entry for "Base year ineligible study" and "Followback study of excluded students.")

Bias (of an estimate): The difference between the expected value of a sample estimate and the corresponding true value for the population.

Bilingual Education: any of various programs in which a student is taught both in English and in the native language so that students keep up in basic subjects while learning English. While there are many varieties of bilingual education, the most commonly encountered is transitional bilingual education. Transitional bilingual education is designed to provide structured English-language instruction and, to the extent necessary to allow a LEP child to achieve competence in English, instruction in the native language of the child, and incorporate the cultural heritage of the child and other children in American society.

Burden: Formally, this is the aggregate hours realistically required for data providers to participate in a data collection. Burden also has a subjective or psychological dimension: the degree to which providing information is regarded as onerous may depend on the salience to the respondent of the questions that are being posed and on other factors such as competing time demands.

BYI: Base Year Ineligible. A baseline (1988) excluded eighth grader. See also entry for Base Year Ineligible Study (above).

CCD: Common Core of Data. Data annually collected from all public schools in the United States by the National Center for Education Statistics.

Ceiling effect: The result of a cognitive test having insufficient numbers of the more difficult items. In a longitudinal study, ceiling effects in the follow-up testings can cause change scores to be artificially constrained for high ability examinees. More information (that is, smaller error of measurement) is obtained with respect to ability level if high ability individuals receive relatively harder items (and if low ability individuals receive proportionately easier items). The matching of item difficulty to a person's ability level yields increased reliability at the extremes of the score distribution where it is most needed for studies of longitudinal change. That is, the measurement problems related to floor and ceiling effects in combination with regression effects found at the extreme score ranges seriously hamper the accuracy of change measures in longitudinal studies. Hence one strategy employed in NELS:88 to minimize ceiling effects was to develop test forms that are "adaptive" to the ability level of the examinee. The multilevel tests used in the first and second follow-ups of NELS:88--with test assignment based on prior test performance--work to minimize the possibility of ceiling effects biasing the estimates of the score gains. (See entry for "Floor effect.")

Cognitive test battery: One of the two parts of the NELS:88 Student Survey (the second part being the student questionnaire). Four achievement areas (mathematics, reading, science, and social studies [history/ citizenship/geography]) were measured.

Cohort: A group of individuals who have a statistical factor in common, for example, year of birth or grade in school or year of high school graduation. NELS:88 embraces three overlapping but distinct nationally-representative grade cohorts: 1987-88 eighth graders, 1989-90 high school sophomores, and 1991-92 high school seniors.

Confidence interval: A sample-based estimate expressed as an interval or range of values within which the true population value is expected to be located (with a specified degree of confidence).

Contextual data: In NELS:88, the primary unit of analysis is the student (or dropout), and information from the other study components, referred to as the contextual data, should be viewed as extensions of the student data--for example, as school administrator, teacher, and parent reports on the student's school learning environment or home situation.

Cross-sectional survey: A cross-sectional design represents events and statuses at a single point in time. For example, a cross-sectional survey may measure the cumulative educational attainment (achievements, attitudes, statuses) of students at a particular stage of schooling (for example, eighth grade, tenth grade, or twelfth grade). In contrast, a longitudinal (or repeated measurement of the same sample units) survey measures the change or growth in educational attainments that occurs over a particular period of schooling. The longitudinal design of NELS:88 generates--by means of sample "freshening"--three representative cross-sections (eighth graders in 1988, high school sophomores in 1990, seniors in 1992) and permits analysis of individual level change over time through longitudinal analysis and of group level and intercohort change through the cross-sectional comparisons. (See entry for "Longitudinal or Panel Survey.")

Dropout: The term is used both to describe an event--leaving school before graduating--and a status --an individual who is not in school and is not a graduate at a defined point in time. The "cohort dropout rate" in NELS:88 is based on measurement of enrollment status of 1988 eighth graders two and four years later (that is, in the spring term of 1990 and the spring term of 1992) and of 1990 sophomores two years later. A respondent who has not graduated from high school or attained an equivalency certificate and who has not attended high school for 20 consecutive days (not counting any excused absences) is considered to

be a dropout. In contrast, transferring schools--for example, from a public to a private school--is not regarded as a dropout event, nor is delayed graduation (as when a student is continuously enrolled but takes an additional year to complete school). A person who drops out of school may later return and graduate: at the time the person left school initially, he or she is called a "dropout," and at the time the person returns to school, he or she is called a "stopout."

ECLS--Early Childhood Longitudinal Study: NCES study now in the design phase scheduled to begin with a kindergarten cohort with a baseline data collection in 1998-99.

Excluded Student: a student deemed ineligible to participate, normally because of a physical or mental disability or language barrier that in the view of the student's school would make participation unduly difficult or impossible. Students excluded in the 1988 NELS:88 baseline are also called "base year ineligible" students (BYIs). The term "excluded student" is also used by NAEP.

ESL: English as a Second Language instruction. Cf. "Bilingual Education" (above).

File: Refers to a data file containing a set of related computerized records.

Floor effect: The result of a cognitive test being too difficult for a large number of the examinees, causing the low ability examinees to receive chance scores on the first testing, and on subsequent testings if the test remains too difficult. Floor effects result in an inability to discriminate among low ability individuals at time one or time two, and there will be no reliable discrimination among examinees with respect to amounts of change. A possible solution, utilized in NELS:88, is to develop test forms that are "adaptive" to the ability level of the examinee, which tends to minimize the possibility of floor effects biasing the estimates of the score gains.

Followback study of excluded students: A continuation in the NELS:88 second follow-up of a special substudy begun in the first follow-up as (see entry for) the base year ineligibles study.

Freshening: A NELS:88 sampling procedure by which high school sophomores were added in the first follow-up who were not in the eighth grade in the U.S. two years before. This process was repeated in the second follow-up, adding high school seniors who were not in the eighth grade in the U.S. four years before, and not in the tenth grade in the U.S. two years before. This process ensured that the *sample* would be representative of the 1992 senior class by allowing 1992 seniors who did not have a chance for selection into the base year (or the first follow-up) sample to have some probability of 1992 selection.

GED test: General Educational Development test. A test administered by the American Council on Education as the basis for awarding a high school equivalent certification.

HS&B: High School and Beyond. The second in the series of longitudinal education studies sponsored by NCES. The HS&B Base Year study surveyed sophomore and senior students in 1980.

IEP: Individualized Education Program in special education for students with a mental or physical disability. Disabilities are defined in terms of the following:

- Specific learning disabilities
- Speech or language impairments

- Mental retardation
- Serious emotional disturbance
- Multiple disabilities
- Hearing impairments
- Orthopedic impairments
- Other health impairments
- Visual impairments
- Autism
- Deaf-blindness
- Traumatic brain injury

IRT: Item Response Theory. IRT models the probability of answering a test item correctly as a mathematical function of proficiency or skill, thus permitting calibration of item parameters and creation of a common scale on which performance can be compared across groups (for example, those who took different versions of a test) or across time (for example, in NELS:88, test results for 1988, 1990 and 1992 have been put on the same scale). IRT estimates achievement level by considering the pattern of right, wrong, and omitted responses on all items administered to an individual student, taking into consideration characteristics of each of the test items, such as their difficulty and the likelihood that they could be guessed correctly by low-ability individuals.

Item nonresponse: The amount of missing information when a valid response to a questionnaire item or variable was expected. (See entry for "Unit-nonresponse.")

LEP: Limited English Proficient. A concept developed to assist in identifying those language-minority students (individuals from non-English language backgrounds) who need language assistance services, in their own language or in English, in the schools. (See entries for "NEP" and "LM.") The Bilingual Education Act, reauthorized in 1988 (PL 100-297), describes a limited English proficient student as one who:

1) meets one or more of the following conditions:

   a) the student was born outside of the United States or the student's native language is not English;

   b) the student comes from an environment where a language other than English is dominant; or

   c) the student is American Indian or Alaskan Native and comes from an environment where a language other than English has had a significant impact on his/her level of English language proficiency; and

2) has sufficient difficulty speaking, reading, writing, or understanding the English language to deny him or her the opportunity to learn successfully in English-only classrooms.

However, LEPs are in fact differently defined in different settings, on the basis of differing methods of determining English proficiency. Some school districts employ English literacy assessments, others rely on teacher reports or home language surveys or formal interviews, or other means to determine who will be classified as an LEP.

LM: Language Minority. LM students come from homes in which a non-English language is spoken. The English language skills of language minority children range from inability to speak English at all to full fluency in English. The NELS:88 base year questionnaires defined LM as follows: A student in whose home a non-English language is typically spoken. Such students may include those whose English is fluent enough to benefit from instruction in academic subjects offered in English as well as students whose English proficiency is limited.

Longitudinal or panel survey: In a longitudinal design, similar measurements--of the same sample of individuals, institutions, households or of some other defined unit--are taken at multiple time points. NELS:88 employs a longitudinal design that follows the same individuals over time, and permits the analysis of individual-level change. (See entry for "Cross-sectional survey.")

Microdata (microrecords): Observations of individual sample members, such as those contained on the NELS:88 data files.

NAEP: The National Assessment of Educational Progress, sponsored by the National Center for Education Statistics (NCES). Since 1969, NAEP has conducted regular assessments of the nation's students in a variety of subjects. Grades 4, 8, and 12 are assessed; a trend sample is maintained as well, comprising three age cohorts: age 9, 13, and 17. NAEP supplies cross-sectional estimates for group-level performance (for example, comparisons of achievement by race/ethnicity, gender, type of community, and region). In contrast, the NELS:88 achievement battery is designed to support the analysis of individual-level cognitive growth over time.

NCEO: The National Center on Educational Outcomes for Students with Disabilities. NCEO was established in 1990 and is supported through a cooperative agreement with OSEP. The center works with state departments of education, national policy-making groups, and others to facilitate the development and use of indicators of educational outcomes for students with disabilities.

NCES: The National Center for Education Statistics, Office of Educational Research and Improvement, of the U.S. Department of Education. This governmental agency is the primary sponsor of NELS:88, and is also the sponsoring agency for (among other studies) NAEP, ECLS, HS&B, and NLS-72.

NELS:88: The National Education Longitudinal Study of 1988. Third in the series of longitudinal education studies sponsored by NCES. The study began in 1988 with the eighth-grade class of that year. The study collected data in 1988, 1990, 1992 and 1994 on student's school experiences, as well as background information from school administrators, teachers and parents; a follow-up is planned for 1998, when most sample members will have been out of high school for six years. The study seeks to learn about students' educational experiences and outcomes from eighth grade through high school and beyond.

NEP: No English Proficiency. A student who does not speak English. (See entry for "LEP.")

NLS-72: The National Longitudinal Study of the High School Class of 1972. This project was the first in the series of longitudinal education studies sponsored by NCES.

NLTS: National Longitudinal Transition Study of Special Education Students.

Nonresponse: (See entry for "Item nonresponse" and "Unit nonresponse.")

Nonsampling error: An error in sample estimates that cannot be attributed to sampling fluctuations. Such errors may arise from many sources including imperfect implementation of sampling procedures, differential unit or item nonresponse across subgroups, bias in estimation, or errors in observation and recording.

NORC: The National Opinion Research Center at The University of Chicago. NORC conducts NELS:88 for the National Center for Education Statistics.

OBEMLA: The Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education. OBEMLA funded a NELS:88 supplement that inquired into the education experiences of students whose native language is other than English.

OCR: Office for Civil Rights, U.S. Department of Education.

OSEP: Office of Special Education Programs, United States Department of Education. OSEP is a unit within OSERS (see below).

OSERS: Office of Special Education and Rehabilitative Services, U.S. Department of Education

Oversampling: disproportionate selection of a sample unit. For example, in the NELS:88 base year, private schools with eighth grades were selected at a higher rate than public schools. Students who were Asian or Hispanic were likewise oversampled so that greater numbers would be available for analysis. The weighting scheme adjusts for differences in the selection probabilities of sample units.

Out-of-sequence: This term means that a student is not in the grade that he/she would be in if progressing with the majority of the grade (or age) cohort through school. For example, most NELS:88 sample members were in the tenth grade in the 1989-90 school year; one would be described as out-of-sequence if found to be in the ninth (or eleventh) grade in the 1989-90 school year.

Performance assessment: a performance-based test is defined by the GAO (1993) as "A test that measures ability by assessing open-ended responses or by asking a person to complete a task. Also known as alternative assessment, constructed response, or task performance, performance-based tests require the respondent to produce a response or demonstrate a skill or procedure. Examples include answering an open-ended question, conversing in a foreign language, solving a mathematical problem while showing all calculations, writing an essay on a given topic, or designing a science experiment."

Population: All individuals in the group to which conclusions from a data collection activity are to be applied. Weighted results of NELS:88 data provide estimates for populations and subgroups.

Poststratification adjustment: A weight adjustment that forces survey estimates to match independent population totals within selected poststrata (adjustment cells).

PPVT: The Peabody Picture Vocabulary Test (Dunn & Dunn, 1981). The revised PPVT (PPVT-R) is currently being succeeded by a newer version, the PPVT-III. The PPVT is a receptive vocabulary test that is normed for ages 2.6 through 40 years. For all items, the subject points to one of four pictures which illustrates a vocabulary item spoken by the examiner.

Precision: The difference between a sample-based estimate and its expected value. Precision is measured by the *sampling error* (or *standard error*) of an estimate.

Probability sample: A sample selected by a method such that each unit has a fixed and determined probability of selection.

QED: Quality Education Data. QED is a commercial firm that publishes national directories of all public and private schools and districts. Its list of schools in the U.S. constituted the sampling frame for the base year, and provided important information on school location, principal's name, minority enrollment, and other characteristics.

Reliability: The consistency in results of a test or measurement including the tendency of the test or measurement to produce the same results when applied twice to some entity or attribute believed not to have changed in the interval between measurements. (For detailed discussion of reliability issues in the NELS:88 achievement battery, see Rock & Pollack, 1995.)

Sample: Subgroup selected from the entire population.

Sampling error: The part of the difference between a value for an entire population and an estimate of that value derived from a probability sample that results from observing only a sample of values.

Sampling variance: A measure of dispersion of values of a statistic that would occur if the survey were repeated a large number of times using the same sample design, instrument and data collection methodology. The square root of the sampling variance is the *standard error*.

Standard deviation: The most widely used measure of dispersion of a frequency distribution. It is equal to the positive square root of the *population variance*.

Standard error: The positive square root of the *sampling variance*. It is a measure of the dispersion of the sampling distribution of a statistic. Standard errors are used to establish *confidence intervals* for the statistics being analyzed.

Student questionnaire: One of the two parts of the student survey (the other part is the cognitive test battery). This instrument contained a locator section for tracing sample members for future waves of NELS:88 and a series of questions about courses taken, hours spent on homework, and perceptions of the school and the home environment. The 1990 and 1992 student questionnaires were available in English and in Spanish.

Survey day: A day chosen by the school during the data collection period when an NORC interviewer and a clerical assistant administered the survey to the school's sample of students. The survey day

session lasted about three hours for the actual data collection, with about thirty minutes each for preparation and clean-up/preparation of completed materials for mailing.

Transfer student: in the NELS:88 base year, some students transferred into a school between student sample selection and the survey day, while others transferred out. Consequently, a sample update was conducted in the base year so that transfers into the school prior to survey day would be represented. After the base year, members of the student sample were followed (though with some subsampling) regardless of the schools they transferred to.

Unit nonresponse: Failure of a survey unit (for example, at the institutional level, a school, or at the individual level, a respondent, such as a student or a teacher) to cooperate or complete survey instrument. Unit nonresponse may be contrasted to item nonresponse, which is the failure of a participating sample member to give a valid response to a particular question on a survey instrument.

Validity: The capacity of an item or measuring instrument to measure what it was designed to measure; stated most often in terms of the correlation between scores in the instrument and measures of performance on some external criterion. *Reliability*, on the other hand, refers to consistency of measurement over time. (See entry for "Reliability.") (For detailed discussion of validity issues in the NELS:88 achievement battery, see Rock & Pollack, 1995.)

Weighted estimates: Estimates from a sample survey in which sample data are statistically weighted (multiplied) by factors reflecting the sample design. The weights (referred to as sampling weights) are typically equal to the reciprocals of the overall selection probabilities, multiplied by a nonresponse or poststratification adjustment. Thus, for example, the 1,035 completed school administrator questionnaires in the NELS:88 base year represent a population of 38,774 schools. Individual completed cases (that is, base year school administrator questionnaires) may "represent" anywhere from a minimum of 1.5 schools to a maximum of 387.3 schools. To take another example, 12,111 base year questionnaire respondents reported themselves to be male, and a slightly greater number (12,244) reported themselves to be female. When these cases are multiplied by the nonresponse-adjusted student weights to yield a weighted percent that reflects the national population of eighth graders, the estimate for males is 50.1 percent of the 1988 eighth-grade cohort while females are estimated to comprise 49.9 percent of the nation's 1988 eighth graders.

Woodcock-Johnson. The Woodcock-Johnson Psycho-Educational Battery Revised (Woodcock & Johnson, 1989) comprises both cognitive and achievement subtests covering the age range from age four through adulthood.

★ U.S. GOVERNMENT PRINTING OFFICE: 1996 – 411 - 108 / 50357

United States
Department of Education
Washington, DC 20208–5651

Official Business
Penalty for Private Use, $300

Postage and Fees Paid
U.S. Department of Education
Permit No. G–17

**Fourth Class Special
Special Handling**